

DELHI COLLEGE OF ENGINEERING



LIBRARY
Kashmiri Gate, Delhi-110006

Accession No. _____

Class No. _____

Book No. _____

PREFACE TO SECOND EDITION

In preparing the present edition the main structure of the book has not been altered, but a number of major or minor changes have been made in every chapter. The purpose of these changes has been either to bring the information up to date or to improve and clarify the treatment. At the request of many teachers who have used the first edition in their classes, an important change has been made in the thermodynamic symbols. It is not certain that the symbols now used in the text are the most satisfactory, but at least they correspond to those most widely encountered in the literature of physical chemistry.

By decreasing the size of some of the figures and using smaller type for the references, economies of space have been effected. Nevertheless, an increase in the already large size of the book has been inevitable. It is hoped, however, that the advantage of a one-volume work will more than offset the disadvantage of its bulk.

The author has been gratified by the friendly comments he has received from readers of the book in all parts of the world, and he wishes to thank his numerous correspondents for their many helpful suggestions. He is unfortunately not able to include the names of all those who have contributed to the changes made in this edition, but mention must be made of Dean Hugh S. Taylor and of Professor Roy F. Newton to whom he is especially indebted. In conclusion, the writer would like to record his appreciation of the help given him by his wife in the reading of proof and in numerous other ways.

S. G.

BERKELEY, CALIF.
March 1946

EXTRACT FROM PREFACE TO FIRST EDITION

In common with other teachers of physical chemistry, the writer has felt for some time the need for a textbook of a standard between the elementary works, which serve as an introduction, and those of a much more advanced character intended for students with a fair knowledge of the subject. It is in an attempt to fill the gap in the literature of physical chemistry that this book has been written. Its object is to take the student with a very elementary knowledge of the subject, and to lead him by easy stages and with the simplest mathematical methods to such an understanding of physical chemistry as will permit him to appreciate the more advanced treatises and the recent journal literature. In an effort to make the book as comprehensive as possible a large number of topics have been dealt with; it is hoped that this will serve the purpose of making the book useful as a work of reference for all chemists, whether they be teachers or students or are engaged in some aspect of chemical industry. Specialists in particular fields may perhaps feel that the subjects in which they are interested have received only superficial treatment; in view of the limitations of space it has clearly been necessary to keep the material within bounds, and the considerable number of references given at the end of each chapter will indicate to the reader where he may find further information. In general, the papers referred to are either of historical interest, contain reviews of the subject under consideration, or deal with the most recent work.

In writing this book the author has tried to give what may be called the "modern point of view" in physical chemistry, instead of grafting, as has often been done, the more recent concepts on to a background of classical ideas. In spite of this, emphasis is laid on the historical development of the various subjects discussed, for in the writer's opinion the historical approach often provides the best way of inculcating the research outlook which should form the basis of all science teaching. Another unusual feature of the book is the description of experimental methods. It is true that physical chemistry involves the theoretical interpretation of experimental observations, but all too frequently the student is quite unaware of the methods used in making these observations. For this reason sections throughout the book have been devoted to the discussion of the fundamental principles concerned and to brief descriptions of the apparatus employed in the measurement of various physical properties. No attempt has been made to give detailed instructions for carrying out experiments, but sufficient is stated for the reader to understand the methods used in physicochemical investigations of various kinds.

It will be noted that about one third of the book is in small type; the material printed in this manner includes the descriptions of experimental

methods, the discussion of more advanced topics and relatively difficult mathematical treatments, all of which might be omitted on a first reading. The reasons for adopting the particular arrangement of the chapters in the book are explained in the introduction, but it is by no means necessary that this order should be followed exactly. The possibility of individual choice in the matter has been kept in mind and frequent cross references are given so as to facilitate a study of the book in any reasonable order.

As a teacher of some experience, the writer is well aware of the great value to be attached to the solution of numerical problems as an aid to the understanding of the principles of physical chemistry. It has become the practice, especially in the more elementary texts, to include a selection of such problems at the end of each chapter; it was felt, however, that for an advanced book, such as the present one, problems of a more general character would be preferable. This requirement is met more adequately by works devoted exclusively to physicochemical problems, and since one or two of these are available, it was considered preferable to devote the limited space available to a selection of references for further reading by the student.

S. G.

PRINCETON, N. J.,
May 1940

CONTENTS

CHAPTER	PAGE
PREFACE.....	v
INTRODUCTION.....	xi
I. ATOMIC STRUCTURE AND ATOMIC SPECTRA.....	1
II. RADIOACTIVITY AND ISOTOPES.....	118
III. FIRST AND SECOND LAWS OF THERMODYNAMICS.....	181
IV. THE GASEOUS STATE.....	244
V. THE SOLID STATE.....	340
VI. CHANGES OF STATE.....	427
VII. THE LIQUID STATE.....	479
VIII. PHYSICAL PROPERTIES AND MOLECULAR STRUCTURE.....	524
IX. DILUTE SOLUTIONS.....	625
X. PHASE EQUILIBRIA.....	693
XI. CHEMICAL EQUILIBRIUM AND FREE ENERGY.....	815
XII. ELECTROCHEMISTRY.....	884
XIII. CHEMICAL KINETICS.....	1044
XIV. SURFACE PHENOMENA.....	1194
INDEX.....	1285

INTRODUCTION

For purposes of study it has been found convenient to divide natural phenomena into two broad classes: one consists of changes of an apparently permanent nature involving the transformation of one form of matter into another, while in the second are included temporary changes, generally resulting from an alteration of external conditions. It is the study of the phenomena in these two categories which constitute the sciences of chemistry and physics, respectively. The distinction between these aspects of the study of nature may be indicated in another way: chemistry may be said to deal with matter and its transformations, whereas physics is concerned with energy and its transformations. It is clearly not possible to draw a sharp distinction between the two points of view, for many problems in both physics and chemistry are concerned with interactions between energy and matter; it is these problems which constitute the fundamental basis of the subject of physical chemistry.

It has been said that chemistry deals with the combination of atoms and physics with the forces between atoms; obviously, atomic combination involves atomic forces, and it is one of the objects of physical chemistry to see how far the chemical interactions observed between atoms and molecules can be interpreted by means of the forces existing within and between atoms. It is reasonable, therefore, that a book on physical chemistry should begin, as does the present one, with a consideration of the structure of the atom. An understanding of this subject requires a knowledge of spectra as well as of quantum theory, and so these topics are presented together with a brief discussion of the fundamentals of wave mechanics. The study of atomic structure leads to the problem of why atoms combine, and this is considered in connection with the electronic theory of valence. For many reasons, historical and fundamental, the development of modern views concerning the structure of the atom is bound up very closely with the subject of radioactivity, and so this provides the material for the next chapter. The study of isotopes, artificial disintegration of atoms and induced radioactivity then follows in a natural sequence.

In addition to the atomic or molecular approach to the problem of chemical reaction, for which the earlier chapters provide the groundwork, there is another method of attack that has proved very valuable: this involves the use of thermodynamics, which deals with energy changes only and is not concerned with the detailed behavior of atoms and molecules. The treatment of the first and second laws of thermodynamics, the former including thermochemistry, provides an adequate basis for a consideration of the physical properties, structure and interconversion of the gaseous, liquid and solid states of matter. The gaseous state is undoubt-

edly the most simple and so this is dealt with first. A discussion of liquids might be expected to follow, but it has become evident in recent years that a great deal can be learnt about liquids from what is already known concerning the structure of solids, and so chapters on the solid state and on the phenomena associated with change of state precede the treatment of liquids, liquid crystals and the vitreous form of matter.

One of the objectives of physical chemistry is to be able to apply measurements of physical properties, such as density, surface tension, refractive index, dielectric constant, magnetism, optical activity, etc., to the elucidation of chemical structure. Much of the knowledge gained in the earlier sections of the book can now be turned to use in this connection, and a chapter on the relation between physical properties and chemical structure shows what has been achieved in this field of investigation.

The treatment hitherto has been concerned almost exclusively with single substances, and now systems consisting of two or more constituents come up for consideration. It is convenient to divide such systems into two groups, namely, dilute solutions and saturated solutions, which are examined from different points of view. The properties of dilute solutions are discussed from the standpoints of vapor pressure, osmotic pressure and related properties, and the question of the value and limitation of these properties for the determination of molecular weights in solution is reviewed. On the other hand, saturated solutions, and systems involving two or more phases in general, are preferentially studied with the aid of the phase rule and equilibrium diagrams, and this is the procedure adopted in the succeeding chapter.

The time is now opportune for an examination of chemical reactions, for the study of their mechanisms and of the influence on them of the physical environment, e.g., temperature, radiation and the presence of catalysts. In a sense it might have been logical to deal with the kinetics, or rates, of chemical reactions before treating equilibrium processes, but much of the information obtained in the study of the latter is utilized in connection with the former, and so the reverse order has been adopted. Further, since a knowledge of certain properties of ions is required for the discussion of reactions occurring in solution, it has been found advisable to interpolate a chapter on electrochemistry. This could not have preceded the treatment of equilibria, as an understanding of this subject is essential for the adequate treatment of the properties of ionic solutions. It may be mentioned, incidentally, that electrochemistry provides an excellent illustration of the statement made above that physical chemistry deals with the interaction between energy and matter, for many of the phenomena considered are the result of the passage of an electric current through solutions of electrolytes.

The book is concluded with a consideration of a number of aspects of the properties of interfaces, including colloidal solutions, since the latter owe their characteristics, to a great extent, to surface effects. Much of the material given could have followed, perhaps with advantage, the

discussion of the three principal states of matter or of ordinary solutions, but as many of the properties of colloidal systems are electrical in nature it was considered advisable to leave the whole subject until after the treatment of electrochemistry.

One of the purposes of this brief introduction, as indicated in the preface, is to explain the reasons for the particular arrangement of the material in this book. On the whole the reader will find it best to adhere to the order given, but it will be observed that the inclusion of cross references makes reasonable changes possible without inconvenience.

CHAPTER I

ATOMIC STRUCTURE AND ATOMIC SPECTRA

THE ELECTRON

Faraday's Laws and the Atom of Electricity.—In the course of his researches on the passage of electric currents through solutions, M. Faraday (1834) discovered that the same quantity of electricity always set free at the electrodes equivalent weights of different substances. Accurate determinations have shown that the liberation of 1 gram equivalent of any substance requires the passage of 96,490 absolute coulombs of electricity, and this quantity, represented by the symbol F , is called a **faraday**. If it is assumed that current is carried through a solution by ions, which are charged particles of matter, it follows that 1 g. equiv. of any ion carries the same amount of electric charge, namely F coulombs. The number of gram equivalents in 1 g. ion is equal to the valence of the latter, by definition of the equivalent, and so it follows from Faraday's law that 1 g. ion of any substance carries nF coulombs, where n is the valence. The number of individual ions in 1 g. ion is N , the Avogadro number,* and so the electric charge associated with every separate ion is nF/N . The importance of this result was realized in a general way by G. Johnstone Stoney (1874) and more clearly by H. von Helmholtz (1881).¹ Since n is an integer, viz., one for a univalent ion, two for a bivalent ion, and so on, it follows that the charge of electricity carried by any single ion is a multiple of the fundamental quantity represented by F/N . In other words, electricity, like matter, is atomic and F/N is the "atom," or unit, of electric charge for which Stoney (1891) proposed the name **electron**. The arguments based on the work of Faraday thus showed that the passage of electricity through solutions involves units of charge associated with matter; a number of important discoveries made towards the end of the 19th century proved the concept to be capable of considerable extension.

The Passage of Electricity through Gases.—Under normal conditions air, and other gases, are very poor conductors of electricity, and a discharge will pass in the form of a spark if a high potential is employed. At reduced pressures, however, conduction occurs more readily and is associated with a number of luminous effects. The phenomena observed in the so-called Geissler, or discharge, tubes are complicated and have been the subject of much investigation by J. Plücker (1858), W. Hittorf (1869), W. Crookes (1879), and others; it is proposed to deal here only

* The number of individual atoms in a g. atom, of ions in a g. ion, or of molecules in a mole is called the Avogadro number.

with those aspects that are relevant to the main problems of physical chemistry. As the gas pressure is gradually diminished it is found that a glow surrounding the cathode, i.e., the negative electrode, detaches itself leaving a space between it and the electrode; this is known as the Crookes dark space, and at sufficiently low pressure, viz., 0.01 to 0.001 mm. of mercury, it fills the whole tube. At this stage there is emitted from the cathode a stream of rays, known as **cathode rays**, which produce a fluorescence on the walls of the discharge tube. The rays travel in straight lines normal to the cathode, as may be shown by the shadows thrown by opaque articles placed in their path; they possess sufficient momentum to cause a small paddle wheel to rotate, and are deflected in electric (E. Goldstein, 1876) and magnetic (J. Plucker, 1858; W. Hittorf, 1869) fields. The latter effects suggest that the rays are associated with an electric charge and J. Perrin (1895) and J. J. Thomson (1897) were able to show that this had a negative sign. The most reasonable interpretation of the facts described is that the cathode rays consist of negatively charged particles expelled from the cathode with a high velocity; J. J. Thomson (1897) proposed to name these particles "corpuscles," but later they were called **electrons**, since they are apparently identical with the unit of electricity, mentioned above, associated with ions in solution. The name electron was meant by G. J. Stoney to refer to the unit of either positive or negative electricity; as used at present, however, it generally implies the unit negative particle, since the corresponding positive particle was not discovered until many years later (p. 20).

Positive Rays.—Before describing in more detail the work on cathode rays, brief mention may be made of other discharge-tube phenomena of importance. In 1886, E. Goldstein found that if a perforated disc was used as cathode, colored rays were seen, provided the pressure was not too low, to emerge on the side opposite to the anode, i.e., the positive electrode. These rays were called **canal rays** since they passed through the holes or channels in the cathode, but they are now referred to as **positive rays**, for they are positively charged and are deflected by a magnetic field in a manner opposite to that to be expected for cathode rays moving in the same direction (W. Wien, 1900). The positive rays are always associated with material particles, and in this respect differ from cathode rays; they also differ in the respect that whereas the latter actually *start* from the cathode, the former appear to originate in the glow surrounding the Crookes dark space. True anode rays have been detected (E. Gehecke and O. Reichenheim, 1906) when using a glass tube containing an alkali halide as anode; the rays emitted consist of positively charged ions of the metal present in the salt. The results obtained from a study of positive and anode rays have had a profound influence on the knowledge of atomic structure, as will be seen later.

X-Rays.—W. C. Röntgen (1895) observed that when cathode rays, produced in a highly evacuated discharge tube, impinged on matter new

radiations, called **X-rays**, were produced. The most convenient method for obtaining them is to place a metal target, called an anti-cathode, in the path of the cathode rays; the X-rays are then emitted from the anti-cathode. These rays were found to be more penetrating than cathode rays, being able, for example, to affect a photographic plate wrapped in black paper; the greater the penetrating power of the rays the greater is said to be their "hardness." X-rays are able to excite fluorescence in certain salts, e.g., zinc silicate containing a trace of manganese, and barium platinocyanide, and to produce ionization in gases so that they are able to conduct electricity. Electric and magnetic fields have no influence on X-rays, and so they evidently do not consist of charged particles; both G. G. Stokes (1898) and G. J. Stoney (1898) suggested that X-rays are electromagnetic waves, similar in character to light waves but of much shorter wave length. This view, now generally accepted, has been confirmed by the fact that X-rays show diffraction phenomena (Chapter V).

The Properties of Cathode Rays: The Ratio of Charge to Mass of the Electron.—The earliest quantitative studies of cathode rays were published almost simultaneously in 1897 by E. Wiechert (Jan.), W. Kaufmann (July), and J. J. Thomson (Oct.). The principle of Thomson's method for the determination of the ratio of charge to mass of the particles assumed to constitute the rays is as follows. Consider a stream of such negatively charged particles, each of charge e and mass m , moving in a straight line with a velocity v , and suppose a magnetic field of strength H is applied at right angles to the direction of motion; the magnetic force exerted on each particle is then Hev . Under the influence of this constant force the particles will describe a circular path of radius r , in which the centrifugal force mv^2/r just balances the magnetic force; it follows, therefore, that

$$Hev = mv^2/r \quad (1)$$

and

$$e/m = v/Hr. \quad (2)$$

The radius of curvature r of the path of a cathode ray stream, when deflected in a field of known strength H , can be measured, but a knowledge of the velocity of the particles is still necessary for the charge to mass ratio, e/m , to be evaluated.

An electric field of strength X acts on a particle of charge e with a force Xe , and if the strengths of the magnetic and electric fields, acting simultaneously but in opposition, are so adjusted that the beam of rays is not deflected from its original direction, it follows that the magnetic and electric forces are equal, so that

$$Hev = Xe, \quad \therefore \quad v = X/H. \quad (3)$$

It is thus possible to determine the velocity of the particles from the ratio of the strengths of the two fields, expressed in appropriate units.

The apparatus used by J. J. Thomson is shown diagrammatically in Fig. 1: the rays emitted from the cathode *C* passed through a hole in the anode *A*, and a narrow beam was picked out by the slit *S*. By connecting the plates *PP'* to a source of high potential an electric field could be applied to the cathode rays, and a magnetic field could be produced with the aid of an electromagnet

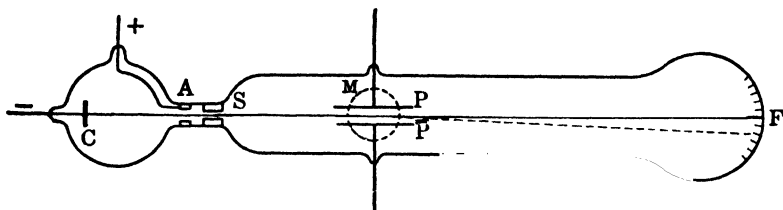


FIG. 1. Determination of the velocity of electrons (J. J. Thomson)

M outside the tube. The deflection of the beam was determined by observation of the luminous spot produced by the rays when striking the wall of the tube; a scale *F* on the outside permitted actual measurement. After noting the position of the undeflected beam the magnetic field was applied, and from the deflection the radius of curvature of the path was calculated; then the electric field was put on and adjusted so as to bring the ray back to its normal position.

From these and other measurements, such as observations of deflection in an electric field only, and determination of the kinetic energy of the particles constituting the cathode rays, the value of e/m was found by the workers mentioned above to be approximately 10^7 electromagnetic (c.g.s.) units per gram. The same result was obtained irrespective of the nature of the gas present in the discharge tube, or of the material of the cathode.

In addition to the method described, involving the use of discharge through gases at high potentials, others are possible for which relatively low voltages, e.g., 100 volts, are adequate. If the cathode is exposed to ultraviolet radiation, quite appreciable currents may be made to flow through a gas at low pressure. This action of the radiation on the cathode material is known as the **photoelectric effect**. Use may also be made of **thermionic effects**; if the cathode is heated to incandescence, for example by means of a subsidiary current, conduction in the gas is also possible. By coating the cathode with certain oxides, e.g., calcium or barium oxide with a trace of cerium oxide, it needs to be heated only to a relatively low temperature before the gas becomes conducting. In both of the cases described it is only the electrode connected to the negative side of the battery, i.e., the cathode, which on heating or exposure to radiation renders the gas conducting; the anode is quite ineffective in this connection. It appears, therefore, that the photoelectric and thermionic effects cause the emission of *negatively* charged carriers of electricity, and experiments of the type described for cathode rays have shown that these have e/m values almost identical with that already given.

The striking constancy of the ratio of charge to mass for the negatively charged particles produced under a variety of conditions suggests that

these particles, namely electrons, are always the same and represent the fundamental carriers of electricity in gases. The fact that these electrons are obtainable from different electrode materials, leads to the important conclusion that they are a universal constituent of all matter. Under the influence of the high voltage in the discharge tube a stream of electrons is sent out from the cathode, and similarly the effect of ultraviolet radiation and of high temperature is to produce emission of electrons which permit electricity to be carried through a gas at low pressure. At higher gas pressures the current in a discharge tube may be carried, to some extent, by gaseous ions, that is by material particles which have acquired an electric charge (see p. 8).

It should be pointed out that electrons carry a negative charge, and so move in a direction opposite to that of the normal flow of "positive current"; thus, the current is said to flow from anode (positive electrode) to the cathode (negative electrode) through the conducting medium, but the electrons move in the opposite direction. This confusion arises because of the early conventions adopted concerning the direction of the flow of current in a voltaic cell.²

Electrical Units.—The electrostatic force F between two charges e and e' placed at a distance r apart is given by $F = ee'/\kappa r^2$, where κ depends on the nature of the medium. Since the dimensions of force are mlt^{-2} (see p. 181), where m represents mass, l is length and t is time, it can be readily shown that the dimensions of electric charge are $m^{1/2}l^{3/2}t^{-1}\kappa^{1/2}$, the dimensions of κ not being known. The strength of an electric current is equal to the rate at which an electric charge moves along a conductor, so that the dimensions of current must be $m^{1/2}l^{3/2}t^{-2}\kappa^{1/2}$. The electromagnetic force between two poles of strength m and m' is given by $F = mm'/\mu r^2$, where μ depends on the medium, and so the dimensions of pole strength must be $m^{1/2}l^{3/2}t^{-1}\mu^{1/2}$. It can be shown that the work done in carrying a magnetic charge m round a closed path carrying current i is equal to $4\pi mi$, so that the product of pole strength and current should have the dimensions of work, i.e., ml^2t^{-2} . According to this argument, therefore, the dimensions of current should be $m^{1/2}l^{3/2}t^{-1}\mu^{-1/2}$. Since the dimensions of current must be the same irrespective of the method of calculation, it follows that

$$m^{1/2}l^{3/2}t^{-2}\kappa^{1/2} = m^{1/2}l^{3/2}t^{-1}\mu^{-1/2}; \quad \therefore \kappa^{1/2}\mu^{1/2} = l^{-1}t = 1/c. \quad (4)$$

The dimensions lt^{-1} are those of a velocity, and hence $\kappa^{1/2}\mu^{1/2}$ for any given medium must be the reciprocal of a velocity; both experimental and theoretical treatment has shown this to be the velocity of light (c).

In practice κ and μ are assumed to be unity in a vacuum, and units are defined accordingly as electrostatic (e.s.) and electromagnetic (e.m.), respectively.* It is obvious, however, from (4) that unit current on the e.s. system cannot be identical with unit current on the e.m. system, for if κ is 1 for a given medium μ cannot be the same. From a consideration of (4) it may be readily deduced that

$$\frac{1 \text{ e.s. unit of current}}{1 \text{ e.m. unit of current}} = \frac{1}{c},$$

* In these systems of units κ and μ , respectively, become dimensionless.

so that if c is 3×10^{10} cm. per sec., it follows that the e.m. unit of current is 3×10^{10} times as large as the e.s. unit in the c.g.s. system.* Since electric charge, or quantity of electricity, is equal to the current strength multiplied by time, it is clear that 1 e.m. unit of charge is also 3×10^{10} as great as an e.s. unit. Further, the product of quantity of electricity and potential is equal to work, and if the same unit, viz., c.g.s., of work or energy is adopted in each case the e.m. unit of potential must be smaller than the e.s. unit in the ratio of 1 to 3×10^{10} . The work done in moving 1 e.s. unit charge against 1 e.s. unit of potential is 1 erg.

In addition to the e.m. and e.s. units, there are the so-called "practical" units, which are of a more convenient size. The practical unit of current is the *ampere*, defined as one-tenth the e.m. (c.g.s.) unit, and the corresponding unit of quantity of electricity, or charge, is one *coulomb*, i.e., 1 amp. sec.; the practical unit of potential is the *volt*, defined as equal to 10^8 e.m. units. Bearing these ratios in mind, and those relating e.m. to e.s. units, it is possible to draw up a table, giving the factors whereby readings in any one unit are to be multiplied for conversion into another unit. Strictly speaking the value 2.9977×10^{10} should be used for the velocity of light, but for simplicity the approximate figure 3×10^{10} is employed.

TABLE 1. CONVERSION FACTORS

	E.M. to E.S. Units	E.M. to Practical Units	E.S. to Practical Units
Current	3×10^{10}	10	0.33×10^{-9}
Quantity or Charge	3×10^{10}	10	0.33×10^{-9}
Potential	0.33×10^{-10}	10^{-8}	300

For many purposes it is convenient to define a unit of energy called the **electron-volt (E.V.)**: it is the energy acquired by an electron, carrying a charge of 4.802×10^{-10} e.s. unit, moving under a fall of potential of 1 volt, i.e., $1/300$ e.s. unit. The work done is, therefore, $4.802 \times 10^{-10}/300$, i.e., 1.601×10^{-12} erg; hence this is the c.g.s. equivalent of 1 E.V. Another important energy unit is the **volt-coulomb**, which is the work done by a flow of 1 coulomb of electricity under 1 volt of potential; from the definitions given above this is clearly 10^7 ergs.

The Mass of an Electron.—According to the most recent work, the ratio of charge to mass of the electron is 1.759×10^7 e.m. units per g., i.e., 1.759×10^8 coulombs per g., irrespective of its origin, provided its velocity is not too high; this may be compared with the corresponding ratio for the positive hydrogen ion in solution. A g. atom of hydrogen weighs 1.008 g. and carries a charge of 96,490 coulombs, so that e/m is 95,700 coulombs per g., which is about $1/1838$ of that of an electron. If the electron had the same mass as a hydrogen atom, the lightest atom known, its charge would be 1838 times as great as that of the hydrogen ion; on the other hand, if it has the same charge, the mass of the electron is only $1/1838$ th part of the mass of a hydrogen atom. It is very prob-

* This is defined as the current which flowing through a wire in the form of an arc 1 cm. long and 1 cm. radius exerts a force of 1 dyne on 1 e.s. unit magnetic pole placed at the center of the arc.

able, as will be seen shortly, that the charge carried by an ion in solution is the same as that of an electron, and so it follows that the mass of an electron is $1/1838$, i.e., 0.000544 , on the ordinary atomic weight scale.

Mass and Velocity.—In stating the value of e/m for an electron, the condition was made that the velocity should not be high. In Thomson's classical experiments the observed speed was about 3×10^9 cm. per sec., that is, about one-tenth of the velocity of light, but the speed of an electron can be varied by changing the potential gradient between anode and cathode. The speed of the electron is proportional to the square root of the potential gradient, and so the value may be readily determined. It has been found experimentally that at very high speeds e/m commences to decrease, and this decrease continues as the velocity of the electron is increased. The change is not due to any alteration in the charge, but to an increase of mass, as required by the theory of relativity. According to this theory, the actual mass m of a particle moving with a velocity v is given by

$$m = m_0 / \sqrt{1 - (v/c)^2}, \quad (5)$$

where m_0 is the mass of the particle at rest, often called the "rest mass," and c is the velocity of light. It follows from (5), therefore, that the effective mass of an electron, and in fact of any particle, should increase with increasing velocity. If v is less than 10^9 cm. per sec. the value of m differs from the rest mass m_0 by not more than 1 per cent, and so the effect of increase of mass is not apparent until the speed approaches that of light. At relatively low velocities the mass of an electron may, therefore, be regarded as constant and almost equal to the rest mass; it is this quantity which is $1/1838$ th of the mass of a hydrogen atom. The evaluation of the mass requires a knowledge of the electronic charge and so will be given later (p. 13).

Mass and Energy.—The relationship between mass and velocity leads to an important result concerning the connection between mass and energy. A force F acting on a body which moves through a distance dx increases the kinetic energy by dE , where dE is equal to Fdx . According to Newton's second law of motion, force is equal to the rate of change of momentum, that is F equals $d(mv)/dt$; hence,

$$dE = Fdx = d(mv)dx/dt. \quad (6)$$

Since dx/dt is the expression for the velocity v of the body, it follows that

$$dE = vd(mv) = v^2dm + mv dv. \quad (7)$$

The relativity equation (5) may be written in the form

$$m^2(c^2 - v^2) = m_0^2c^2, \quad (5a)$$

and differentiation gives

$$(c^2 - v^2)dm - mv dv = 0,$$

and by introducing (7) it follows that

$$dE = c^2dm. \quad (8)$$

It is seen from this equation that any increase in kinetic energy must be accompanied by a proportionate increase of mass; this conclusion can be extended to other forms of energy, and it follows that in general

$$\Delta E = c^2 \Delta m. \quad (9)$$

A change of energy ΔE is, therefore, associated with a change of mass equal to $\Delta E/c^2$, where c is the velocity of light (3×10^{10} cm. per sec.). The law of conservation of mass is, consequently, only strictly true if there is no energy change in the process considered. For chemical reactions the difference of mass corresponding to the heat change involved is so small as to be outside the limits of measurement; it is for this reason that the most accurate determinations have led to the apparent verification of the law of conservation of mass. It can be shown from (9) that 100,000 calories, i.e., 418×10^{10} ergs, is equivalent to no more than 4.6×10^{-9} g.; thus,

$$\begin{aligned} \Delta m &= \Delta E/c^2 = 418 \times 10^{10} / (3 \times 10^{10})^2 \\ &= 4.6 \times 10^{-9} \text{ g.}, \end{aligned}$$

and so could not normally be detected. It will be seen later, however, that the relationship between mass and energy leads to appreciable effects in certain cases, although these do not involve chemical reactions in the ordinary sense (see p. 165).

The Charge of an Electron.—The direct measurement of the electronic charge has not been found possible, but by making certain reasonable assumptions its value may be obtained. When a gas is rendered electrically conducting, for example, by exposure to cathode rays, to X-rays, or to the radiations emitted by radioactive substances (p. 119), it is said to be ionized, since there are now charged particles, i.e., ions, present in the gas capable of carrying a current of electricity under the influence of an applied potential. The ions probably consist of molecules of gas which, by the action of the ionizing agent, have been made to gain or lose one or more electrons, thus acquiring a negative or positive charge, respectively, which is in magnitude an integral multiple of the charge of a single electron. It will be evident that, since the molecules of gas are initially electrically neutral, the loss of an electron means the acquisition of a positive charge equal to that of an electron. If this view of gaseous ionization is correct, it follows that the minimum charge or, more correctly, the highest common factor of the charges, carried by gaseous ions is equal to the electronic charge.

The earliest determination of the charge of gaseous ions was made by J. S. Townsend (1897), who employed ionized gases liberated by electrolysis of aqueous solutions with large currents; the gas obtained from sulfuric acid solution was found to contain positive ions, whereas that from aqueous potassium hydroxide was negatively charged. When these gases were bubbled through water they formed a dense cloud, apparently resulting from the condensation of moisture on the gaseous ions, consisting

of minute drops of water carrying electric charges. The total charge carried by the cloud in a given volume of gas was measured by means of a quadrant electrometer, and the weight of the cloud was obtained by absorbing the water in concentrated sulfuric acid. In order to estimate the number of droplets in the cloud, its rate of fall under gravity was observed and Stokes's law (p. 497) applied in the form $v = 2gr^2\rho/9\eta$, where v is the uniform rate of fall of a spherical drop of radius r , ρ is its density, g the acceleration due to gravity, and η is the coefficient of viscosity of the medium in which the droplets fall. By this means the radius, and hence the average volume and mass, of a drop of water, assumed to be spherical, was obtained, and since the total mass is known the number of droplets in a given volume of gas could be calculated. The number of drops and their total charge having been determined, the mean charge per gaseous ion was evaluated on the assumption that every ion had a droplet of water condensed on it. The result obtained was approximately 3×10^{-10} e.s. unit per ion for both positive and negative ions; if each molecule of gas loses or gains only one electron on ionization, this represents the magnitude of the electronic charge.

It is of interest to compare this value with the unit charge carried by an electrolytic ion in solution; this, as already shown, is equal to F/N , where F is 96,490 coulombs and N , the Avogadro number, may be taken as approximately 6×10^{23} (see p. 263), so that the unit charge is about 1.6×10^{-19} coulombs, or 4.8×10^{-10} e.s. unit, which is in striking agreement with that evaluated for gaseous ions. Further confirmation of this equality, and incidentally of the views, already expressed (p. 1) concerning the atomic structure of electricity, was obtained by J. S. Townsend (1900), when he showed, by a method not involving any knowledge of the Avogadro number, that ions in gases and solution carried the same charge. The coefficient of diffusion (D) of a gaseous ion is related to its mobility (u), that is, its velocity under a potential gradient of one volt per cm., by the equation

$$u = neD/P, \quad (10)$$

where e is the charge on an ion, P is the pressure of the gas, and n the number of molecules per cc. of gas. The quantities u and D were measured for gases ionized by X-rays, so that the quantity ne could be evaluated; it was found to be approximately 1.3×10^{10} e.s. units for the ions in various gases. The analogous quantity ne for ions in solution was obtained in the following manner: the charge (e) on each univalent ion, as already seen, is F/N , and the number of molecules (n) in 1 cc. of a gas is $N/22,400$, since N is the number in 1 mole occupying 22,400 cc. at ordinary temperature and pressure. It follows, therefore, that ne is equal to $F/22,400$, i.e., 96,490/22,400 coulombs, which is 1.3×10^{10} e.s. units. Although the exact agreement with the value for gaseous ions is to some extent fortuitous, there is little doubt of the main contention that the unit charge is the same for ions of different types.

Measurements of the charge carried by ions produced in gases by exposure to X-rays and to radioactive materials were made by J. J. Thomson (1898, 1903), who employed the discovery of C. T. R. Wilson

(1897) that sudden expansion of a moisture-laden atmosphere causes droplets of water to condense on ionized molecules. The total charge of the cloud was evaluated by conductivity measurements, and the size of the drops was determined, as in Townsend's work, by using Stokes's law; an indirect method was used for estimating the total weight of the cloud in a given volume of gas. The mean charge of a gaseous ion was found in this way to be about 6.6×10^{-10} and 3.4×10^{-10} e.s. unit for ionization by X-rays and radioactive materials, respectively. The discrepancy between these two values is to be attributed to experimental errors, as well as to the failure of Stokes's law to apply exactly, and to the incorrectness of the assumption that each droplet of water is associated with one unit charge.

In a further development, which has provided the basis for later methods of estimating the electronic charge, H. A. Wilson (1903) avoided the necessity of determining the total charge carried by the cloud. The rate of fall (v_1) of the cloud was measured under the influence of gravity alone, when the force acting was m_ag , where m_a is the mass of the droplet; another measurement (v_2) was then made when the fall of the drops was accelerated by the application of an electric field of strength X , the total downward force being then $m_ag + Xc$, if c is the charge carried by each droplet. It follows, therefore, that

$$\frac{m_ag}{m_ag + Xc} = \frac{v_1}{v_2}, \quad (11)$$

the uniform velocities being proportional to the forces acting on the drops. The average mass m_a of a drop was evaluated as in previous methods by application of Stokes's law.³

In 1909, F. Ehrenhaft and R. A. Millikan, independently, made an advance in technique by observing single droplets, instead of the whole cloud, and later (1913), in a further improvement, Millikan made use of oil, instead of water, for attachment to the gaseous ions produced by means of X-rays. In this way errors due to evaporation, and hence to change of weight of the drops, were avoided.

The apparatus employed by Millikan is shown diagrammatically in Fig. 2: the vessel *A*, immersed in a thermostat, contained air at a pressure regulated by a pump and manometer. It was fitted with an atomizer *B* whereby a finely divided spray of oil could be introduced. Of the two plates, *C* and *D*, the upper, which had a number of small holes, as at *E*, was connected to one terminal of a high voltage battery; the other terminal and the lower plate (*D*) were grounded. When a drop of oil was observed to pass through the hole *E*, and enter the space between the plates, the hole was closed and the air ionized by exposure to X-rays passing through the window *W*₁. The oil drop attached itself to a gaseous ion, thus acquiring a charge; it could consequently be made to move up or down, according to the sign of the charge and the direction of the electric field which could be reversed at will. A powerful beam of light from an arc lamp entered the window *W*₂ illuminating the oil drop, and observations on it

were made with a short-focus telescope placed at an angle to the beam. The time taken by the drop to fall under the influence of gravity alone through a distance corresponding to that between two sets of cross wires in the telescope was measured; then, an electric field of 5,000 to 10,000 volts was applied so as to make the drop move upwards, in opposition to gravity, and the time again

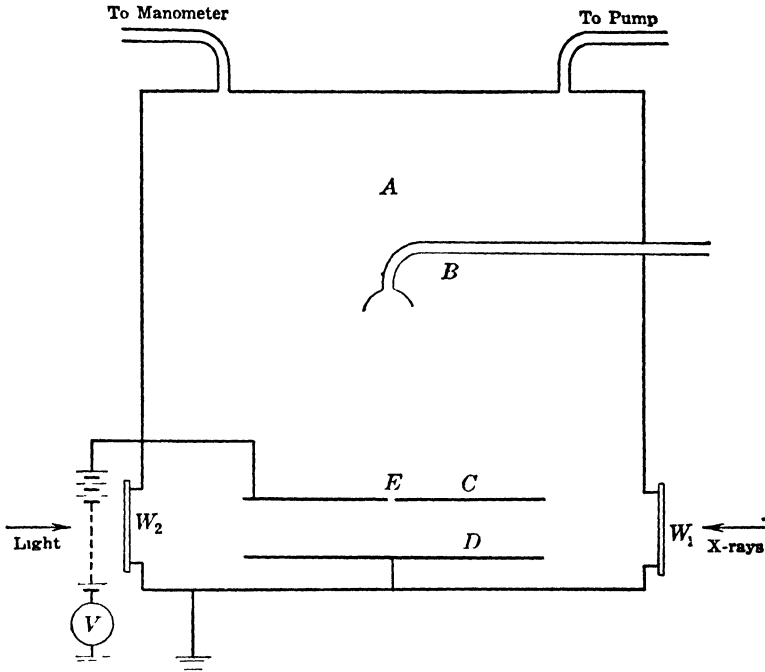


FIG. 2. Determination of the electronic charge (Millikan)

noted. The battery was switched off, the plates short-circuited, and the times of fall under gravity alone and of rise when the same electric field was applied were again determined. This procedure was repeated several times in succession with the same oil drop. Measurements were also made with different drops and electric fields of different strengths.

If v_1 is the rate of fall of the drop under gravity, and v_2 is the rate of upward movement when an electric field of strength X acts in opposition to gravity, then

$$\frac{v_1}{v_2} = \frac{m_d g}{X e_n - m_d g}, \quad (12)$$

where m_d is the mass of the oil drop and e_n the charge it carries. Since the weight of the drop remains constant, v_1 should also be constant, but v_2 might vary with the charge e_n carried if the number of gaseous ions attached to the oil drop altered in successive journeys. The results

showed that such an alteration frequently occurred: the rate of fall of an oil drop under gravity was always constant, but the rate of upward movement, when an electric field was acting, was not the same each time. For example, in one experiment the average time taken to fall between the cross wires in the telescope was 13.60 sec., whereas successive times of rise were 12.5, 12.4, 21.8, 34.8, 84.5, 85.5, 34.6, 34.8, 16.0, 16.0, 34.8, 34.6 and 21.9 sec., indicating that in the various upward journeys the droplet had captured different numbers of ions. Knowing the actual distance corresponding to that between the cross wires in the telescope, the values of v_1 and v_2 were determined, and hence e_n , the corresponding charge, calculated from (12), if the weight m_d of the drop was known. To obtain this quantity use was made of a form of Stokes's equation which included an allowance for the fact that the gaseous medium, that is, the air, consists of molecules whose mean free path is comparable with the radius of the oil drop. The rate of fall v_1 under the influence of gravity of a spherical drop of radius r is then given by the expression

$$v_1 = \frac{2gr^2}{9\eta} (\rho_d - \rho_a) \left(1 + A \frac{\lambda}{r} \right), \quad (13)$$

where ρ_d and ρ_a are the densities of the drop and of air, respectively, η is the viscosity of air, λ the mean free path of the molecules under the experimental conditions, and A is an empirical constant. From the radius of the drop, calculated from (13), and the density of the oil, its mass was obtained, so that all the information required for the calculation of e_n was available. If the variation of upward velocity under the influence of the electric field is due to differences in the charge carried by the oil drop, and if this charge is always an integral multiple of a single unit, namely the electron, the values of e_n should always be multiples of a definite highest common factor representing the unit electronic charge. The results obtained amply confirmed this expectation, thus providing strong evidence for the atomic nature of electricity.

Using the best available datum for the viscosity of air in (13), namely 1822.7×10^{-7} c.g.s. unit at 23°C ., Millikan concluded (1917) that the value of the unit electric charge, which was presumably identical with that of the electron, was $(4.774 \pm 0.005) \times 10^{-10}$ e.s. unit. This result was accepted for several years as one of exceptional accuracy, but later, chiefly as a consequence of precise determinations of the Avogadro number by methods involving the use of X-rays, as explained below, it became evident that Millikan's value for the electronic charge was too low, mainly because of an error in the viscosity of air employed by him. Newer values for this quantity applied to Millikan's observations indicate that the electronic charge is close to 4.800×10^{-10} e.s. unit.

The original oil drop experiments were liable to various errors, but many of these have now been eliminated especially in the form of apparatus used by I. Hopper and T. H. Laby (1941). The electric field is horizontal, instead of vertical, and the rate of fall of the relatively large

v , its momentum is mv , and so

$$mv = h/\lambda. \quad (19)$$

Representing the momentum by p , it follows that

$$\lambda = h/mv = h/p. \quad (20)$$

This is the fundamental equation of de Broglie's theory, for it gives the wave length λ of the hypothetical **matter waves** associated with material particles.

The velocity v of the particle is called the "group velocity" of the matter waves, but this differs from the velocity u of the waves themselves, that is, their "phase velocity." Since the wave length is equal to the velocity divided by the frequency, it follows that for matter waves, $\lambda = u/\nu$, and hence $\nu = u/\lambda$; by the quantum theory, therefore, the energy of the particle, i.e., $h\nu$, would be hu/λ . According to the relativity theory of the relationship between mass and energy, that is by (9),

$$hu/\lambda = mc^2, \quad (21)$$

and introducing the value of λ , from (20), there results

$$uv = c^2. \quad (22)$$

The product of the group and phase velocities of matter waves is thus equal to the square of the velocity of light. Since no material particle can travel with a velocity greater than that of light, for otherwise its mass would become infinite [cf. equation (5)], v must be less than c , and consequently the phase velocity of material waves must always be greater than the velocity of light. The physical significance of this fact is not altogether clear, but it shows that matter waves differ fundamentally from the electromagnetic waves constituting ordinary radiation.

The theoretical significance of the wave theory of matter as applied to electrons will be discussed later (p. 73), but it is appropriate to refer here to an important practical aspect, namely, the discovery that electrons, and other material particles, are capable of being diffracted in a manner exactly analogous to light and X-rays. It follows, therefore, that electrons must be associated in some manner with wave properties, and hence the theoretical views of de Broglie, and their development by E. Schrödinger and others, find immediate justification.

The velocity v of an electron depends on the potential drop through which it passes. If V is the potential and e the electronic charge, then the energy of motion will be Ve , and this may be equated to the kinetic energy, $\frac{1}{2}mv^2$, so that

$$v = \sqrt{2Ve/m}, \quad (23)$$

and since by (20), $\lambda = h/mv$, it follows that

$$\lambda = h/\sqrt{2meV}. \quad (24)$$

From the known values of the Planck constant h , and the charge e and mass m of the electron, it is found that h/\sqrt{me} is practically equal to 10^{-8}

in c.g.s. units, so that

$$\lambda = 1/\sqrt{2V} \times 10^{-8} \text{ cm.} \quad (25)$$

The potential V here is in e.s. units, but it can be expressed in volts by introducing the conversion factor given on page 6; thus,

$$\lambda = \sqrt{\frac{150}{V}} \times 10^{-8} \text{ cm.,} \quad (26)$$

where V is now the potential fall in volts by which the electrons are accelerated. From (26) it is seen that the effective wave length of an electron depends on the accelerating potential, but for values of the order of 100 to 1000 volts the wave length is about 10^{-8} cm., which is similar to that of X-rays.

Provided the potential is low, e.g. less than 10,000 volts, the mass of the electron may be taken as the rest mass, as was done in obtaining (25), but at high accelerating voltages allowance must be made for the change of mass corresponding to the high velocity, by means of (5). Neglecting small quantities this leads to the result

$$\lambda = \sqrt{\frac{150}{V}} \cdot \frac{1}{1 + 4.95 \times 10^{-7} V} \times 10^{-8} \text{ cm.,} \quad (27)$$

the term additional to that in (26) being generally called the relativity correction.

It has been known for several years that crystals can act as diffraction gratings for X-rays, since the spacing of atoms and molecules in a crystal is of the same order as the wave length of the rays; layers of similar atoms or molecules thus act as reflecting planes and produce diffraction effects. It was first suggested by W. Elsasser (1925) that evidence of the wave nature of electrons might be obtained in an analogous manner, and the observations of C. J. Davisson and L. H. Germer (1927), following earlier ones of C. J. Davisson and C. H. Kunsman (1923), proved this to be the case. A beam of electrons, from a heated tungsten filament A (Fig. 3), passed through a known fall of potential, and then impinged on a single crystal of nickel B from which they were reflected into the opening of a Faraday cage C , the inner cylinder being connected to a galvanometer. The angle of incidence of the electron beam on the crystal being fixed,

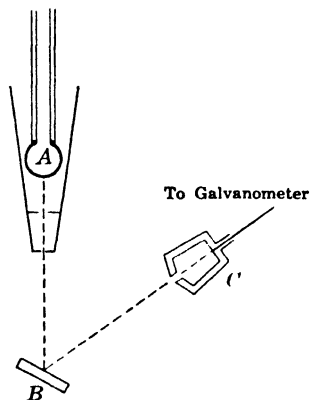


FIG. 3. Diffraction of electrons by a crystal (Davisson and Germer)

the accelerating potential was varied gradually and the intensity of the reflected electron stream determined by the deflection of the galvanometer. The variation of intensity was not regular, and when plotted against the

square root of the accelerating potential a series of almost equally spaced maxima were obtained (Fig. 4). This is exactly the type of behavior to be expected for a wave motion, such as X-rays, reflected from a crystal acting as a diffracting grating, as the wave length is increased; the reflection maxima should be equally spaced when plotted against the reciprocal of the wave length. The results are, therefore, in agreement with (24) which requires the length of the "electron waves" to be inversely proportional to the square root of the applied potential.

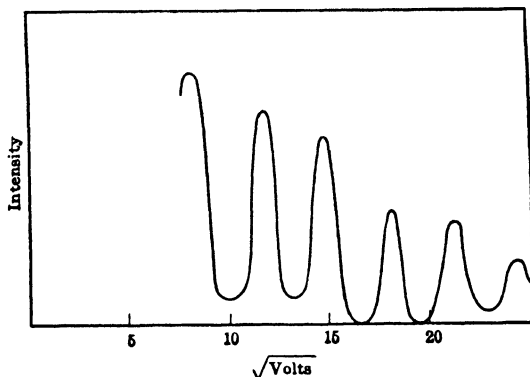


FIG. 4. Intensity of electron diffraction and accelerating potential

In another series of experiments Davisson and Germer cut a single crystal of nickel so as to expose a (111) plane (see p. 352), and an electron beam was allowed to impinge on it at normal incidence. The crystal could be rotated and the position of the Faraday cylinder altered so as to determine the intensities of reflection at different angles. Instead of general scattering, as might be expected for particles, the electrons were found to be scattered most strongly in certain definite directions related to the three-fold symmetry of the nickel crystal. For some positions, with particular voltages, e.g., in a direction at 50° from the incident beam and with a potential of 54 volts, especially intense scattering was observed. In this particular orientation the grating constant is known from X-ray measurements to be 2.15\AA ,* and hence the electron wave length is $2.15 \sin 50^\circ$ (see p. 357), that is 1.65\AA ; the value calculated from (26) for 54-volt electrons is 1.67\AA .

Further evidence for the existence of electron waves was obtained independently by G. P. Thomson (1927) who passed a stream of accelerated electrons, using a potential of 20,000 to 50,000 volts, through a very thin film, about 10^{-6} cm. thickness, of metal, and then allowed it to fall on a photographic plate. On development this showed a series of concentric rings round a central spot representing the direction of the beam. This diffraction pattern suggests that wave properties are associated with electrons, and if the interpretation of the results is correct, then the

* The symbol \AA . refers to the Ångström unit which is 10^{-8} cm.

radius r of any one ring should be proportional to the wave length λ of the electron waves employed, that is r/λ should be constant. Since the value of λ will vary with the applied voltage, according to (27), it follows that the radius r of a given diffraction ring should be related to the applied voltage V by

$$r(1 + 4.95 \times 10^{-7}) \sqrt{V} = \text{constant.} \quad (28)$$

The results of Thomson's observations over a considerable voltage range were in harmony with this equation. It is of interest to record that the diffraction rings could be distorted by means of a magnet, showing they were actually produced by electrons and not by radiation of the type of X-rays which might have been present.

Still another method for demonstrating the wave properties of electrons was devised by E. Rupp (1928): he showed that an electron beam suffered diffraction from an ordinary ruled line grating using extremely small grazing angles, viz., about 0.001° , to compensate for the low wave length (cf. p. 13). From a knowledge of the grating constant and the positions of the diffraction maxima, it was possible to calculate the wave length of the electron waves for various accelerating potentials. These were found to be 1.5 Å., 1.0 Å. and 0.7 Å., for potentials of 70, 150 and 310 volts, respectively, in satisfactory agreement with the theoretical requirements of (26). That the diffraction was produced by electrons was again shown by displacement of the diffraction pattern by a magnetic field.

Apart from the great theoretical significance of the results described, the property possessed by electrons of being capable of diffraction has had a number of practical applications. One of these is the electron microscope which makes possible the observation of particles approaching molecular dimensions. From the purely physicochemical point of view, the most important has probably been that connected with the diffraction by molecules of gases and vapors (p. 591); the study of the surface films of metals by means of electrons has also given results of considerable interest.⁶

Correlation of Wave and Particle Properties: The Uncertainty Principle.—The problem of the correlation of wave and particle properties is one which concerns not only electrons, but also light, and even atoms, for O. Stern and his collaborators (1930, 1932) and E. Rupp (1932) showed that beams of hydrogen and helium were capable of producing diffraction patterns. If a close examination is made into the significance of the description "particle," it can be shown that experiments are designed with the object of obtaining information concerning the position of the electron. On the other hand, if the experimental arrangement employed virtually amounts to the determination of the momentum (or velocity), then it becomes necessary to use the "wave" description, the connection between the momentum and wave length being given by (20). According to the **uncertainty principle**, put forward by W. Heisenberg (1927), and now widely accepted as one of the fundamental bases of

modern physics, *the simultaneous determination of velocity, or any related property, e.g., energy or momentum, and position is impossible.* It follows, therefore, that an electron, for example, cannot *simultaneously* exhibit both particle and wave properties. The uncertainty principle requires that if Δx is the uncertainty in determining the position of an electron and Δp the uncertainty in measuring momentum, then

$$\Delta x \times \Delta p \approx h, \quad (29)$$

where h is the Planck constant. When performing experiments designed for the determination of the velocity of an electron, Δp is made as small as possible so that Δx must be relatively large, and the position of the electron is not known, neither does it matter in the experiment. For example, measurement of the velocity (or momentum) of the electron involves the study of the wave properties and the use of (20); the particle aspect, whereby the position is determined, has then no significance. Similarly, when investigating the position of electrons a knowledge of the momentum is not necessary, so that it is the particle aspect which is important and the wave property does not arise. The two properties of wave and particle must, therefore, be regarded as complementary and not mutually exclusive. For most purposes in chemistry, it is still adequate to regard the electron as a particle, but to be strictly correct the wave properties must be taken into consideration under appropriate conditions, as will be seen subsequently.

In conclusion, it may be emphasized that the wave aspect of electrons, and of matter generally, does not mean that there are *actual* waves associated with material particles. It is more reasonable to consider, as will be shown in relation to the theory of the structure of the atom (p. 76), that the wave properties imply that the probability of finding an electron, or other particle, at any given point is determined by a mathematical relationship analogous to that whereby wave motion can be described. Further than this it is not possible to go at present, and in fact many scientists find it sufficient to admit the existence of particle and wave properties without further inquiry into their physical significance.

Positive Particles: The Proton and the Positron.—The discovery of the unit of negative electricity, free from matter, in the form of the electron naturally led to attempts to discover the corresponding unit of positive charge. Careful study of the positive rays produced in discharge tubes, however, showed the lightest positively charged particle to have a mass almost equal to that of the hydrogen atom, or about 1837 times as heavy as the electron. This particle is probably a positive hydrogen ion, that is a hydrogen atom which has lost an electron; it is known as a **proton**, in accordance with the suggestion of E. Rutherford (1920), from the Greek word (*protos*) meaning "first."

* It should be noted that the proton is the *gaseous* hydrogen ion H^+ and is not identical with the hydrogen ion in solution; the latter is a combination of a proton with at least one molecule of solvent, e.g., H_3O^+ in aqueous solution (see p. 898).

In spite of repeated failures to detect a positively charged particle free from ordinary matter, the relativistic wave theory of P.A.M. Dirac (1928) required the existence of positive electrons corresponding to those of negative charge, and their discovery was reported by C. D. Anderson in 1932. The result was confirmed almost immediately by P. M. S. Blackett and G. P. S. Occhialini (1933) and strangely enough, in view of its apparently elusive nature, several other sources of the positive electron were found within a short time.

The Wilson cloud-track apparatus which played an important part in the discovery of the positive electron has been of great service in the study of atomic structure generally: it is based on the principle already mentioned, that drops of moisture tend to condense on gaseous ions. A chamber, fitted with a piston, contains dust-free air saturated with water vapor; if the air is suddenly expanded the temperature falls and condensation of the water will occur. When a charged particle passes through the air it ionizes the molecules in its path, and since the drops of water tend to condense on these ions, the path of the particle will appear as a faint line of mist, often known as a "fog track" or "cloud track." By suitable methods of illumination the track can be made clearly visible on a dark background and photographed; by making two simultaneous exposures in directions at right angles, the exact path of the ionizing particle may be determined. The cloud-track apparatus was first devised by C. T. R. Wilson (1911), and has since been greatly improved. In some modern forms of the instrument the particle is itself made, by means of a relay, to operate the mechanism for expanding the gas, thus insuring that the condensation of moisture will occur immediately following the passage of the charged particle through the cloud chamber.

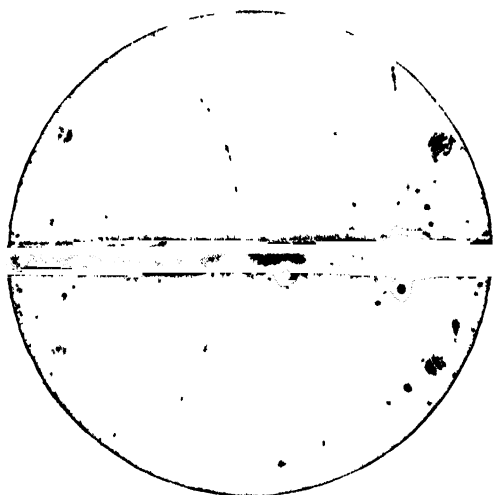


FIG. 5. Cloud-track photograph of positron traveling upward and passing through a 6 mm. lead plate (Anderson)

While studying the properties of cosmic rays, a form of extremely penetrating radiation apparently originating in interstellar space, by means of a vertical Wilson cloud-track apparatus arranged between the poles of a very powerful electromagnet, Anderson observed a track which, by its direction of curvature in the magnetic field, was evidently produced by a positive particle (Fig. 5). The length of the path after passage through a lead plate, however, was at least ten times that possible

for a proton moving in a path of the same curvature, and so it must have resulted from a particle of smaller mass. An examination of the thickness

of the cloud track suggested that the new positive particle had a mass of the same order as that of an electron, and hence evidence was at last obtained of the long sought positive electron. This particle is called the **positron**. The corresponding term **negatron** was suggested for the familiar negative electron, but this is rarely used; the latter is still generally referred to as the electron, without qualification. Anderson's observations, in America, were confirmed shortly afterward in England, and the existence of the positron is now definitely established, especially as it can be obtained in ways other than the action of cosmic rays on matter, for example, when the very penetrating radiations, the γ -rays, produced from the radioactive element thorium-C'' are absorbed by lead, aluminum or other metal. The production of a positron by the action of cosmic or other rays on matter is invariably accompanied by the formation of a negatron, so that cloud-chamber photographs in a magnetic field frequently show two tracks deflected in opposite directions but emerging from the same point. There are, however, generally more negative than positive tracks, since negative electrons are present in atoms and are readily ejected from them.

Mass and Charge of the Positron.—When γ -rays from thorium-C'' are absorbed by a metal the ejected positive particles have a maximum energy, estimated from their deflection in a magnetic field, of 1.6×10^6 electron-volts (e.v.); this is also the maximum value of the energy of positron-negatron pairs observed. Let m_+ and m_- be the masses of the positron and negatron respectively, then by (9) the energy required to produce simultaneously a pair of these particles is $(m_+ + m_-)c^2$. If E is the energy of the γ -ray photon, which on absorption produced the positive-negative electron pair, then the excess energy remaining on the particles must be $E - (m_+ + m_-)c^2$, and this will be equal to the maximum energy which can be acquired by any single particle, or by any combination of positron and electron. As already mentioned, this quantity has been found experimentally to be 1.6×10^6 e.v. The energy of the γ -ray photon (E) is known from its penetrating power to be 2.6×10^6 e.v., and so it follows that

$$\begin{aligned} 2.6 \times 10^6 - (m_+ + m_-)c^2 &= 1.6 \times 10^6; \\ \therefore (m_+ + m_-)c^2 &= 1.0 \times 10^6. \end{aligned}$$

Since m_- , the mass of the electron, is 9×10^{-28} gm. (p. 14), and c , the velocity of light, is 3×10^{10} cm. per sec., m_-c^2 is 8.1×10^{-7} erg, which is equivalent to 0.5×10^6 e.v., so that m_+c^2 must have approximately the same value. It is seen, therefore, that the masses of the positive and negative electrons must be very nearly, if not exactly, identical.

A comparison of e/m values of positive and negative electrons, emitted simultaneously from an artificially radioactive form of copper (p. 167), made under the same conditions, shows that the charge to mass ratios of the two particles do not differ by more than 2 per cent (A. H. Spees and C. T. Zahn, 1940). Since the masses have been shown to be virtually identical, it is evident that the charges must also be the same. The

positron is consequently to be regarded, as far as mass and charge are concerned, as the exact electrical opposite of the familiar negative electron.

Destruction of Positrons.—The positron has only a very short life, probably of the order of 10^{-8} sec.; when its energy has fallen to a relatively low value it suffers annihilation on encountering an electron, and since the negative particles are of frequent occurrence, positrons have only a short period of existence after their formation. It is for this reason that the positive electron had proved so difficult to detect. The masses of both positive and negative particles are electromagnetic in origin, so that when the two combine the result is annihilation accompanied by the production of the equivalent amount of energy, namely $2 \times 0.5 \times 10^6$ e.v. If this energy appears as two photons of radiation moving in opposite directions so as to conserve momentum, each will carry 0.5×10^6 e.v. of energy; by equating this to $h\nu$, according to the quantum theory, and utilizing the known value of h , the wave length of the emitted radiations should be about 2.5×10^{-10} cm. Penetrating γ -rays of this type have in fact been observed to accompany the destruction of positrons.⁷

THE STRUCTURE OF THE ATOM

Introduction.—Until the beginning of the present century no definite ideas concerning the internal structure of atoms had been proposed, and one of the earliest suggestions was made by J. J. Thomson (1904). Since rays consisting of negative electrons having identical properties were obtained from various forms of matter under a variety of conditions (p. 4), it was reasonable to suppose that the electron was an essential constituent of matter, and Thomson considered the atom to be made up of a uniform sphere of positive electricity in which were embedded a number of electrons so that the whole was electrically neutral. The mass of the atom was believed to be due to the electrons only, so that even the lightest atom, namely hydrogen, would need to contain 1838 electrons. This theory encounters so many difficulties that it is unnecessary to consider it further; it is to be regarded, therefore, as purely of historical interest, although it is important in so far as it emphasized the universality of the electron. Another suggestion, made at about the same time by P. Lenard (1903), was based on the observation that swift cathode rays could penetrate thin films of aluminum, and consequently must have passed right through a number of atoms without being deflected. This led Lenard to believe that the greater part of an atom consisted of empty space, and that the material part was made up of neutral doublets, called "dynamids," each consisting, apparently, of a proton and an electron. Only when passing close to a dynamid was it supposed that an electron from the cathode ray stream was deflected appreciably from its course through matter.

The Nuclear Theory of the Atom.—The theory of atomic structure which has proved the most fruitful is the **nuclear theory** proposed by E. Rutherford (1911). It is, in a sense, a more precise development of the general views of Lenard, and of H. Nagaoka (1904), who considered the positive charge of an atom to be concentrated at a central point which was surrounded by negative particles. It will be seen in the next chapter

(p. 119) that certain radioactive elements emit what are known as α -particles; these have a mass four times that of the hydrogen atom and carry two positive charges, and are, in fact, doubly charged positive helium ions. When these α -particles pass through a thin metallic foil most of them suffer no appreciable deflection, but a small proportion are scattered through large angles (H. Geiger and E. Marsden, 1909). For instance, with a gold foil 0.0004 mm. thick, one α -particle in 20,000 was deflected through 90° or more. This scattering must be due to encounters of the α -particles with the atoms of the metal foil, and Rutherford showed that the results could not be explained by ascribing the large deflection to the combined effect of a series of successive small deflections by a number of atoms. He considered, therefore, that the large-angle scattering resulted from a *single* encounter between a metal atom and an α -particle, and showed, on the basis of this view, that the observations could be explained by assuming the whole of the positive charge associated with an atom to be concentrated at a very small center or **nucleus**. This nucleus should be surrounded, at a relatively large distance, by the necessary number of negative electrons, often called the **extra-nuclear electrons**, to give a neutral atom. A fast-moving α -particle will not be deflected from its course by an electron, since the latter has a much smaller mass, and so the majority will pass through an atom without appreciable scattering, but the occasional α -particle which happens to enter the powerful electric field in the vicinity of the minute positively charged nucleus will be violently deflected.⁸

By assuming Coulomb's inverse square law to apply to the interaction between the nucleus of an atom and the positively charged α -particle, the path of the latter can be shown to be a hyperbola with the nucleus N at its external focus (Fig. 6), and Rutherford was then able to deduce an equation for the

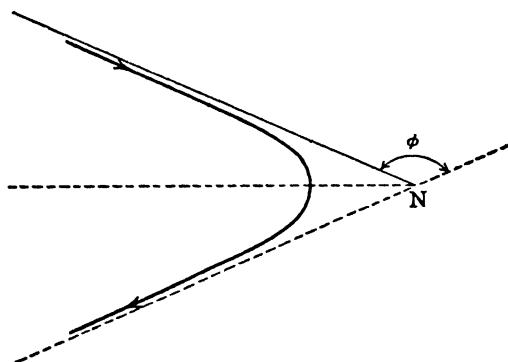


FIG. 6. Scattering of α -particles by an atomic nucleus

extent of scattering in different directions. The fraction $f(\phi)$ of the total number of α -particles scattered over unit area at a distance r from the nucleus, in a direction making an angle ϕ with the original direction of incidence, is

given by

$$f(\phi) = nt \left(\frac{Ze^2}{mv^2r} \right)^2 \text{cosec}^4 \phi/2, \quad (30)$$

where n is the number of atoms per unit volume of scattering material, and t is its thickness; Z is the number of unit charges on the nucleus, and e is the unit charge, i.e., the electronic charge; m is the mass and v the velocity of the α -particle. The general accuracy of (30) was confirmed by the work of H. Geiger (1911) and of H. Geiger and E. Marsden (1913), thus providing strong support for the nuclear atom.

The scattering measurements indicated that the number of charges on the nucleus (Z) was approximately equal to half the atomic weight of the atom, and as the number of electrons surrounding the nucleus must equal the nuclear charge, to maintain neutrality, it follows that this number, in contrast to Thomson's theory, should be relatively small. Since the mass of the electron is also small, it is clearly necessary to postulate as an essential part of the nuclear theory that nearly the whole of the mass of the atom is concentrated on the very small positively charged nucleus.

Dimensions of the Nucleus.—An estimate of the size of the nucleus may be made as follows. The charge on the nucleus is Ze , and that on the α -particle is $2e$, the latter carrying two unit charges, and so, if Coulomb's law applies, the repulsive force is $2Ze^2/r^2$ at a distance r from the nucleus; the corresponding potential energy will be $2Zv^2/r$ (cf. p. 46). The kinetic energy of the α -particle is $\frac{1}{2}mv^2$, and as a rough approximation it may be supposed that the particle commences to reverse its direction when this energy is equal to the potential energy due to repulsion; the distance r at which this occurs, given by

$$\frac{1}{2}mv^2 = \frac{2Ze^2}{r}, \quad (31)$$

may be taken as the effective radius of the nucleus. The mass m of the α -particle is four times that of the hydrogen atom, which is 1838 times that of the electron, so that m is $4 \times 1838 \times 9 \times 10^{-28}$ g.; the velocity v of an α -particle is approximately 2×10^9 cm. per sec. (p. 122). Taking Z as about 10, for an element of atomic weight approximately 20, and the electronic charge e as 4.8×10^{-10} c.s. unit, the value of r is found to be roughly 3×10^{-13} cm. Other calculations made by Rutherford, involving estimates of the smallest distance of approach between an atomic nucleus and an α -particle at which Coulomb's law still applies, indicate that nuclear radii are all of the order of 10^{-12} to 10^{-13} cm. Since the radius of an atom has been found from many measurements to be about 10^{-8} cm., it is clear that an atom must have a relatively "empty" structure, and the observations on the passage of electrons and α -particles can be readily understood. From a knowledge of the radius of an electron (p. 14) the actual volume of matter, that is, of nucleus and electrons,

present in an atom can be calculated to be only about 10^{-12} to 10^{-15} of the effective atomic volume.

The Nuclear Charge.—From the work on the scattering of α -particles it appeared that the nuclear charge was about half the atomic weight, and in 1913 A. van den Broek ⁹ made the important suggestion that the number of charges on the nucleus (Z) is equal to the ordinal number of the element in the periodic table, allowance being made for vacant spaces. In other words, the nuclear charge of an element is equal to its atomic number. This suggestion is in harmony with the results of α -ray scattering, because for many elements, especially the lighter ones, the atomic number is almost exactly half the atomic weight. More precise evidence was obtained at a later date, but in the meantime the classical work of H. G. J. Moseley (1913) on X-rays showed clearly that the atomic number of an element was a property of fundamental importance.

Frequencies of Characteristic X-Rays.—As a general rule the radiations emitted by the anticathode of an X-ray tube are heterogeneous, covering a range of frequencies; if these rays fall on various materials, secondary X-rays are produced with frequencies characteristic of the atoms present in the material. The same characteristic X-rays can be obtained, under suitable conditions, by using an anticathode made of, or containing, the particular element and passing the resultant X-ray beam through a screen of the same material; the screen absorbs extraneous radiation but permits the passage of the characteristic frequencies. The work of C. G. Barkla (1911) showed that the characteristic X-radiations of an element consisted of two rays, or groups of rays, one of which had a much greater penetrating power, that is to say, one was much harder, than the other. The two series were called K - and L -rays, respectively, the former being the harder. The rays of both series become harder as the atomic weight of the element producing them increases, and at the same time the K -radiation becomes more difficult to excite.

With the object of investigating the characteristic X-rays emitted from an anticathode, either made of the element being studied or coated with one of its compounds, H. G. J. Moseley (1913) used a crystal of potassium ferrocyanide as a diffraction grating and allowed the resulting rays to fall on a photographic plate. From the position of the line the frequency of the characteristic radiation could be calculated. In Fig. 7 are seen the photographs for the series of elements from calcium to zinc (brass), with the exception of scandium, from which it is evident that the characteristic frequencies change regularly with the atomic number. The frequency ν of the line belonging to any particular X-ray series was found to be quantitatively related to the atomic number Z of the element by the expression

$$\sqrt{\nu} = a(Z - \sigma), \quad (32)$$

where a is a proportionality constant, and σ is the same for all the lines in a given series; thus for the K_{α} -line, that is, the strongest line in the

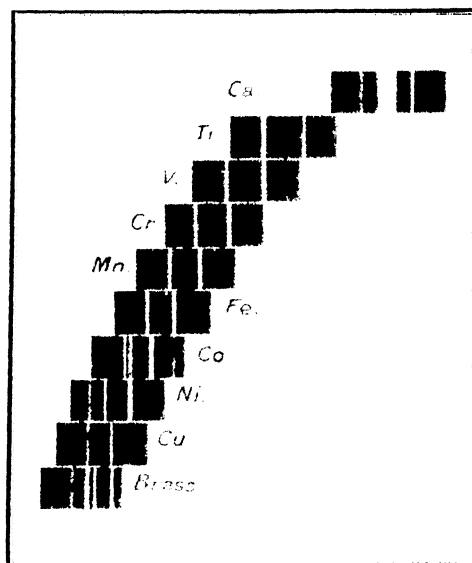


FIG. 7. Characteristic X-rays (Moseley)

K -group, σ is 1.0, and for the strongest line in the L -series for various elements, i.e., for the L_{α} -lines, σ is 7.4. If the square root of the frequency of the characteristic line of a given series is plotted against the atomic number of the element producing the X-rays, a straight line should be obtained, and Moseley found this relationship to hold with fair accuracy for the K -lines given by all elements which could be studied conveniently from aluminum ($Z = 13$) to silver ($Z = 47$), and for the L -series from zirconium ($Z = 40$) to gold ($Z = 79$); the results are shown in Fig. 8.

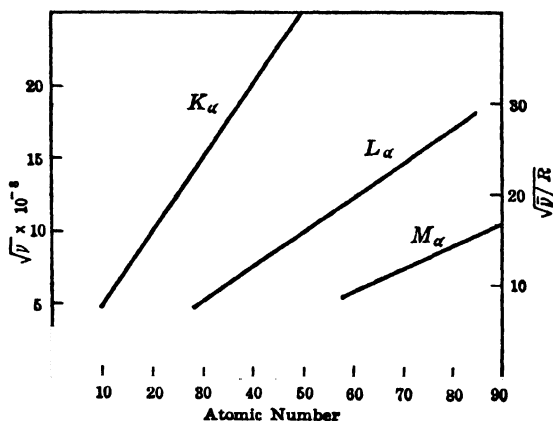


FIG. 8. Dependence of frequency of characteristic X-rays on atomic number

When plotted against the atomic *weights* of the elements, no such regularity was observed. This fact brought out with exceptional clarity the importance of the atomic number of an element, rather than of its atomic weight, and Moseley concluded that there is some property of the atom, either the atomic number or a quantity closely related to it, which is fundamental for each atom and probably represents the magnitude of the positive charge on its nucleus. Later studies have definitely confirmed Moseley's general conclusions; although many other groups of characteristic X-rays, e.g., *M*-, *N*-, and *O*-series, have been discovered, and each is more complicated than was originally realized, the broad relationship between frequency and atomic number is nearly always maintained. In some series, however, abnormalities have been observed for certain elements, and these apparent discrepancies are of great significance in connection with the elucidation of the details of atomic structure (see p. 87).

Applications of Characteristic X-Rays.—Before proceeding to give further evidence for the suggestion that the atomic number of an element is equal to its nuclear charge, brief mention must be made of other important aspects of Moseley's work. Since the frequency of the characteristic X-ray line in a given series is quantitatively related to the atomic number, it is clearly possible to determine the value of the latter if the former has been measured. It is well known that there are three pairs of elements in the periodic classification, viz., argon and potassium; cobalt and nickel; and tellurium and iodine, for which it has been found necessary to invert the usual order of atomic weights in order to comply with the chemical requirements of the table. The X-ray measurements have definitely justified the inversion (see Fig. 7 for cobalt and nickel), indicating that the atomic number, i.e., the order in the periodic table, and not the atomic weight, is the essential property of an element.

A study of the characteristic X-rays has shown that the atomic numbers of molybdenum and ruthenium are 42 and 44 respectively, so that there must be an element between them in the periodic table. It has been claimed that this element, known as masurium, was discovered and identified by its characteristic X-rays. The existence of the element rhenium, atomic number 75, was proved in the same manner (W. Noddack, *et al.*, 1925). It is true that ordinary chemical considerations had suggested that the spaces in the periodic table corresponding to these elements should be kept vacant, but the same certainty did not apply to the group of elements between lanthanum and tantalum, since there was no direct method of finding how many rare-earth elements existed. By means of the characteristic X-rays, however, the atomic numbers of lanthanum and tantalum were found to be 57 and 73 respectively, and so there could only be fifteen in between. Examination of the known rare-earth elements showed that only two spaces, for atomic numbers 61 and 72, remained to be filled. Both missing elements were discovered later, the former was called illinium (B. S. Hopkins, *et al.*, 1925), and the latter—not a rare-earth element, but a homologue of zirconium—was given the name hafnium (D. Coster and G. Hevesy, 1922). In both cases the X-ray method was used for identification purposes. In skilled hands the study of the characteristic X-rays provides an unequivocal method of analysis which does not require separation of the particular element to be identified, provided it is present in appreciable amount. The frequencies of the lines for any element can be calculated, and if they appear

in the X-ray spectrum obtained when a given material is placed on the anti-cathode, then that element must be present in the material; if it is absent, however, a gap will occur in the position the lines should occupy.¹⁰

Direct Determination of Nuclear Charge.—A direct determination of the nuclear charges of the elements copper, silver and platinum has been made by J. Chadwick (1920) using a form of (30) relating the scattering of α -particles to the nuclear charge. Since the number of particles scattered is normally a very small proportion of the total, a special device was employed to increase this fraction. The scattering foil was used in the form of a ring shown in section at AA in Fig. 9, the source of α -par-

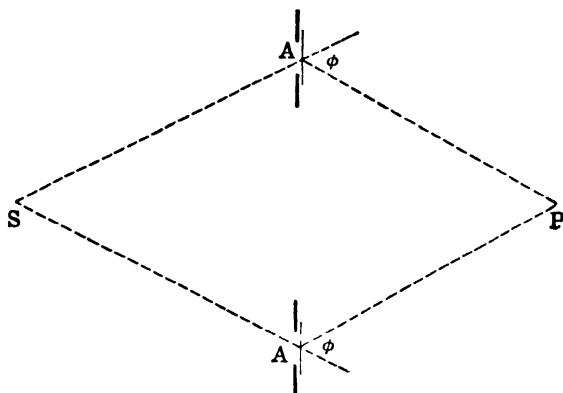


FIG. 9. Determination of nuclear charge

ticles (*S*) being to one side and a small zinc sulfide screen (*P*), for counting the scattered particles by the scintillations they produce (p. 426), on the other. By this arrangement all particles scattered through a definite solid angle were "focussed" on the screen; the proportion of the original number striking the screen was thus about one in a thousand, instead of about one-tenth of this amount in the usual way. The number of incident particles, about 20,000 per min., was estimated by limiting the field of vision with a rotating disc containing a slit of known area compared with that of the whole disc; the total number could then be obtained by simple proportion. The scattered particles struck the zinc sulfide screen at a rate of only 20 or 30 per min., and so the scintillations could be counted directly, without reduction. Working in this manner Chadwick estimated the nuclear charges of copper, silver and platinum to be 29.3, 46.3 and 77.4 units, respectively, with an accuracy between 1 and 2 per cent. These figures are in excellent agreement with the respective atomic numbers, viz., 29, 47 and 78.¹¹

Constitution of the Nucleus: The Neutron.—It can now be regarded as established that the nuclear charge is equal to the atomic number, and so the latter must also give the number of electrons outside the nucleus necessary to produce a neutral atom; these contribute only a very small

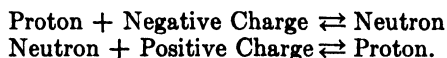
fraction of the total mass, and so the mass of an atom may be assumed to be concentrated on the nucleus. This conclusion was in harmony with the failure, before 1932, to discover a unit positive charge free from appreciable mass, and hence it seemed probable that atomic nuclei were made up of protons, thus accounting for both their charge and mass. The proton has a mass of approximately unity, on the ordinary atomic weight scale, and so the number of protons in a given nucleus should be equal to the atomic weight A of the element. The net nuclear charge is, however, the same as the atomic number Z , which is less than the atomic weight, and so it was suggested that, in addition to protons, the nucleus contained $A - Z$ electrons; these would contribute a negligible amount to the total mass, but would make the *net* positive charge (Z) on the nucleus equal to the atomic number, although the *total* number of positive charges (A) was equal to the atomic weight. The concept of nuclear electrons appeared to receive support from the fact that certain radioactive elements emit electrons, probably from their nuclei, but there was always considerable difficulty in explaining the stability of such a closely packed system of protons and electrons.

In 1920, E. Rutherford had suggested that there might exist particles, which he called **neutrons**, having no charge but with unit mass; for several years, however, there was no experimental evidence of their existence. In the course of an investigation of the effect of α -particles on atomic nuclei W. Bothe and H. Becker (1930) found that certain light atoms, especially beryllium, when bombarded by α -particles emitted a highly penetrating type of radiation not previously encountered. It was shown to be capable of ejecting protons from hydrogen, helium, carbon, air and other substances, and from an analysis of the speeds and masses of the particles concerned J. Chadwick (1932) concluded that mere radiation could not have been responsible for the observed disintegrations. He put forward the suggestion, therefore, that the so-called radiation consisted of particles having the mass of a proton but no charge, and consequently were the neutrons postulated by Rutherford. Further investigation has confirmed this view, and the neutron is now definitely established as a particle capable of independent existence and with interesting properties (see Chapter II). Its mass has been determined with considerable accuracy (p. 166) and found to be 1.00866 on the chemical atomic weight scale, or 1.00893 on the physical ($^{16}\text{O} = 16.0000$) scale (see p. 150).

The part that the neutron might play in nuclear structure was soon realized by W. Heisenberg (1932) and E. Majorana (1933). If a nucleus contains Z protons and $A - Z$ neutrons, the total mass of the atom will equal the atomic weight A , since the neutron and proton have almost the same mass, and the nuclear charge will be equal to Z , the atomic number, as required. For many reasons this view is more satisfactory than the older one which postulated the presence of electrons in the nucleus: one of the most important is that the stability of the nucleus is less difficult to understand. If atomic nuclei consist essentially of neutrons and

protons the atomic weights of all elements should be either exactly or very close to whole numbers. This is probably strictly true, and the apparent deviations observed, e.g., 35.46 for chlorine, are due to the fact that many elements actually consist of mixtures of atoms of different atomic weight, e.g., 35 and 37 for chlorine. This matter will be understood more clearly after the concept of "isotopes" has been discussed (p. 150).¹²

The Meson.—Since the proton and the neutron have almost identical masses, it is probable that they differ merely by a unit electric charge. Thus, the transfer of a negative charge from a neutron to a proton, or of a positive charge in the opposite direction, would result in an exchange of the identities of the two particles, viz.,



Such a possibility would result in an attraction between the neutron and proton as a consequence of what is known as wave mechanical "exchange" or "resonance" energy, to which reference will be made later (p. 80). In this way an explanation could be developed to account for the stability of atomic nuclei. In the original treatment of this subject it was considered that the unit charge by which the neutron and proton differed was the electron or the positron, but this view led to certain results that did not agree with observation. In order to overcome the difficulty, H. Yukawa (1935) found it necessary to postulate the existence of particles of a new type, carrying a unit positive or negative charge, but having a mass of the order of a hundred times that of the electron. These particles were required to decay very rapidly, with an average life of about 10^{-6} sec. In 1937, particles possessing just such properties were discovered in cosmic rays, by means of cloud-track photographs (S. H. Neddermeyer and C. D. Anderson, 1937). Because of their mass, intermediate between that of the electron or positron and that of the proton or neutron, they have been called **mesons** (Greek: *intermediate*). Both positive and negative mesons have been detected in cosmic rays, and their masses are apparently 150 to 220 times that of the electron. It is not yet certain whether this variation is real or whether it is due to experimental discrepancies. It may be mentioned that the existence of a neutral meson, the **neutretto**, has been postulated to account for the attractive forces between protons and protons, and between neutrons and neutrons; such a particle would prove difficult to detect.¹³

The Alpha-Particle in Nuclear Structure.—The atomic number of hydrogen is unity, and the normal atom has, therefore, one electron; if this becomes detached the resulting positively charged particle consists of a nucleus only. Since the proton is probably a positively charged hydrogen atom (p. 19), it appears that the proton is identical with a bare hydrogen nucleus. The α -particle has a mass of four units and, as already stated, there is good reason to believe that it is a helium atom with a double positive charge (see p. 120). The atomic number of helium is two, and so the normal atom has two electrons; if both are lost the result is a bare helium nucleus, with a mass of four, having two positive charges. The α -particle is thus a helium nucleus, and according to the theory of nuclear structure it should consist of two protons, i.e., hydrogen nuclei, and two neutrons. If this is the case the mass of the α -particle should

be 4.0330, made up as follows:

$$\left. \begin{array}{l} \text{mass of two protons} = 2 \times 1.0076 \\ \text{mass of two neutrons} = 2 \times 1.0089 \end{array} \right\} = 4.0330,$$

whereas the actual value is 4.0028,* that is, 0.0302 mass units less than expected. This discrepancy does not necessarily imply that the views concerning the constitution of the α -particle are incorrect, but more probably that in its formation from two protons and two neutrons there is a loss of mass which appears as energy. The energy equivalent of the mass is obtained by multiplying by the square of the velocity of light [equation (7)], so that the amount liberated in the formation of a helium nucleus is

$$E = 0.0302 \times (3 \times 10^{10})^2 = 0.272 \times 10^{20} \text{ ergs per g. atom,}$$

which is equivalent to about 6.6×10^{11} calories. This enormous amount of energy would have to be supplied to 4 g. of helium in order to disintegrate the nuclei into protons and neutrons. It is clear, therefore, that the helium nucleus must be a very stable structure. The fact that helium nuclei, i.e., α -particles, are emitted spontaneously from the nuclei of certain radioactive elements (p. 119) suggests that the latter actually contain helium nuclei as secondary units of structure, and this is in harmony with the marked stability, as just shown, of α -particles. Further evidence for this contention is that the lighter elements having atomic weights which are multiples of four, and so presumably have nuclei made up of α -particles only, are very stable. The only light atoms, for example, which cannot be disintegrated by bombardment with fast α -particles are hydrogen, helium, carbon and oxygen, the atomic weights of the last three being 1×4 , 3×4 and 4×4 , respectively (see p. 161). It may be significant, also, that the atomic weights of the two most common elements, viz., oxygen and silicon, are multiples of four. It is probable, therefore, that protons, neutrons and α -particles are the structural units of atomic nuclei: further aspects of the question of nuclear structure will be considered subsequently (p. 137).

THE QUANTUM THEORY

The Emission and Absorption of Radiation.—Before proceeding with the main subject of this chapter it is necessary to digress so that the fundamentals of the quantum theory of radiation may be introduced. The energy of the radiation emitted by a black body, i.e., a perfect radiator, is not uniform but has a maximum value at a definite wave length which is inversely proportional to the absolute temperature; further, the total energy emitted by the black body varies as the fourth power of the temperature. Attempts to account for these facts had been made by W. Wien (1896) and by Lord Rayleigh (1900), but the equation of the former held only at low temperatures and short wave lengths, whereas that of the latter was applicable only at high temperatures and long wave lengths. The deduction of Rayleigh had been based on the assumption that the radiation was emitted by oscillators, and that the principle of the equipartition of energy (p. 329), according to which the

* These masses are all based on the physical atomic weight scale (p. 150).

mean energy of an oscillator is equal to kT , was applicable.* In 1900, M. Planck, however, discarded this principle and made the entirely new suggestion that an oscillator is unable to absorb or emit energy continuously, but only in integral multiples of a definite amount depending on the frequency of the oscillator. In other words, the energy of a body is not continuously variable, but can consist only of a definite whole number of **quanta**, and that the energy can only be taken up or given out in such **quanta**; this is the basis of the **quantum theory** which has found application in many fields of physics and chemistry. According to Planck, the energy of the quantum (ϵ) is given by

$$\epsilon = h\nu, \quad (33)$$

where ν is the frequency of the oscillator and h is a constant, known as the **Planck constant**; the latter is sometimes called the **action constant** because it has the dimensions of energy \times time, i.e., "action." Equation (33), of which frequent use will be made, is the fundamental equation of the quantum theory.

Mean Energy of an Oscillator.—If N_0 is the number of oscillators in the state of lowest, i.e., zero-point, energy (cf. p. 157), then according to the classical distribution law (p. 272), the number possessing vibrational energy ϵ in excess of the zero-point value will be $N_0 e^{-\epsilon/kT}$. If N is the total number of oscillators, possessing different amounts of energy, $\epsilon_0, \epsilon_1, \epsilon_2$, etc., then this number must equal the sum of the terms representing the numbers with the various amounts of energy; that is,

$$N = N_0 e^{-\epsilon_0/kT} + N_0 e^{-\epsilon_1/kT} + N_0 e^{-\epsilon_2/kT} + N_0 e^{-\epsilon_3/kT} + \dots \quad (34)$$

According to the quantum theory, ϵ can only be a whole number, or zero, of quanta $h\nu$, so that the possible values are 0, $h\nu$, $2h\nu$, $3h\nu$, etc., corresponding to $\epsilon_0, \epsilon_1, \epsilon_2, \epsilon_3$, etc., in (34); making this substitution, it follows that

$$N = N_0 + N_0 e^{-h\nu/kT} + N_0 e^{-2h\nu/kT} + N_0 e^{-3h\nu/kT} + \dots, \quad (35)$$

$$= N_0 \sum_{n=0}^{\infty} e^{-nh\nu/kT}. \quad (36)$$

The total energy E of the oscillators is obtained by multiplying the number in each energy state by the energy of that state, and adding all the products; thus,

$$E = \epsilon_0 N_0 + \epsilon_1 N_0 e^{-\epsilon_1/kT} + \epsilon_2 N_0 e^{-\epsilon_2/kT} + \dots \quad (37)$$

$$= 0 + h\nu N_0 e^{-h\nu/kT} + 2h\nu N_0 e^{-2h\nu/kT} + \dots \quad (38)$$

$$= N_0 \sum_{n=0}^{\infty} nh\nu e^{-nh\nu/kT}. \quad (39)$$

* The constant k is known as the Boltzmann constant, and is the gas constant per single molecule, i.e., R/N , where R is the usual gas constant per mole and N is the Avogadro number (see p. 264).

The average energy $\bar{\epsilon}$ of each oscillator is then equal to E/N ; hence,

$$\bar{\epsilon} = \frac{E}{N} = \frac{h\nu \sum_{n=0}^{\infty} n e^{-nx}}{\sum_{n=0}^{\infty} e^{-nx}}, \quad (40)$$

where x is $h\nu/kT$. The numerator and denominator can be expressed in the form of a series, and if this is done it is found that

$$\bar{\epsilon} = \frac{h\nu}{e^x - 1} = \frac{h\nu}{e^{h\nu/kT} - 1}, \quad (41)$$

which gives the mean energy of an oscillator in excess of the zero-point value. By means of this relationship Planck was able to deduce an equation, connecting the frequency with the energy of the radiation emitted by a black body, in excellent agreement with experiment; the value of h derived from the data was about 6.5×10^{-27} erg sec. It will be observed that when $h\nu/kT$ is small, i.e., for low frequencies (long wave lengths) or high temperatures, (41) reduces to $\bar{\epsilon} = kT$, as used by Rayleigh; hence his equation is applicable only under these conditions. Planck was also able to show that at short wave lengths and low temperatures the expression for the energy of radiation reduced to the same form as that proposed by Wien.

The Photon.—The success of the quantum theory in this connection led A. Einstein (1905) to suggest that radiation is not only absorbed and emitted in quanta, but that it is actually propagated through space, with the velocity of light, in definite quanta or **photons**, equal in energy to $h\nu$, where ν is the vibration frequency of the radiation. Although this represented a break with the traditional view that radiation consists of a wave motion, it has been possible to bring about a reconciliation by postulating that radiations partake of the nature of both waves and particles, but that the different properties are manifest under different conditions (cf. p. 19). By means of this modification of the quantum theory, Einstein deduced the equation

$$h\nu = \frac{1}{2}mv^2 + P \quad (42)$$

for the photoelectric effect; $h\nu$, the energy of the photon of incident radiation, is equal to the sum of $\frac{1}{2}mv^2$, i.e., the kinetic energy of the emitted electron, and the energy P required to release the electron from the metal, i.e., the product of the **work function** of the metal and the electronic charge. It is apparent from (42) that unless the frequency of the incident radiation exceeds a definite minimum value, so that $h\nu$ is greater than P , there can be no photoelectric emission of electrons. This minimum, or threshold, frequency has been found experimentally and may be designated by ν_0 , so that

$$h\nu = \frac{1}{2}mv^2 + h\nu_0 \quad \text{or} \quad \frac{1}{2}mv^2 = h(\nu - \nu_0). \quad (43)$$

This form of the Einstein equation was tested by R. A. Millikan (1916), the kinetic energies of the photoelectrons, for various frequencies of light,

being determined by finding the potential difference necessary to bring them to rest. The value of h found in this manner was 6.57×10^{-27} erg sec., in striking agreement with that derived from the distribution of energy in black body radiation.

In the photoelectric effect the energy of the photon is converted into that of an electron; in the production of X-rays the reverse is the case, and the study of X-rays offers another possibility for the calculation of the Planck constant. The bombardment of a metal by electrons produces X-rays, but no rays are emitted until the electrons acquire a minimum energy by passage through a difference of potential. If V is the applied potential and e is the electronic charge, the energy of the electron is Ve , and this may, according to the quantum theory, be equated to $h\nu$, where ν is the frequency of the X-rays emitted. Since not all the electrons acquire the maximum possible energy Ve , not all the rays have the same frequency; ν must therefore represent the maximum frequency, so that $Ve = h\nu_{\text{max}}$. By means of this relationship W. Duane and R. L. Hunt (1915) obtained a value of h almost identical with those already given. On the basis of the new figures for the electronic charge (p.13) and mass, the Planck constant is now accepted as $(6.624 \pm 0.002) \times 10^{-27}$ erg sec.¹⁴

The Compton Effect.—Important evidence for the particle, or photon, nature of radiation is provided by the discovery of A. H. Compton (1923) that if monochromatic X-rays are allowed to fall on a scattering material of low atomic weight, e.g., carbon, the scattered X-rays contain, in addition to those having the incident wave length, others of a somewhat longer wave length. Since the scattering is actually produced by electrons, it appears that the interaction between X-rays and electrons results in an increase of wave length of the former. By assuming the X-rays to consist of particles possessing energy $h\nu$, and momentum $h\nu/c$, where c is the velocity of light, it can be calculated, by ordinary considerations of the conservation of energy and of momentum, that the increase $\Delta\lambda$ in the wave length should be

$$\Delta\lambda = \frac{2h}{mc} \sin^2 \frac{\theta}{2}, \quad (44)$$

where m is the rest mass of the electron and θ is the angle between the incident and scattered X-rays. According to (44), therefore, the Compton shift of wave length should be independent of the wave length of the incident X-rays; this result has been confirmed, and the value of $\Delta\lambda$ is in close agreement with that calculated from the known magnitudes of m , c and h .

It is of interest to note that the Compton effect provides an excellent illustration of the uncertainty principle (p. 19). Theoretically, it might be possible to determine the position and momentum of an electron by using X-rays as a source of "illumination." The change in wave length of the X-rays, and hence in the momentum of the associated photons, means that there will be a change in the momentum of the electron. Consequently, the magnitude of the latter cannot be determined with certainty.

ATOMIC SPECTRA

Atomic Spectra.—Up to the present the consideration of the extra-nuclear electrons has been restricted to an indication that their number must equal the atomic number of the element, and to a mention of the fact they form a relatively open structure about the central nucleus. The question of their arrangement must now be discussed in more detail, and in this connection valuable information has been obtained from an examination of both optical and X-ray spectra, the former term being used to describe spectra in the ultraviolet, visible and infrared regions.

When a solid is heated to incandescence it emits a more or less continuous spectrum, but gases and vapors under the same conditions, when examined spectroscopically, show a series, often very complicated, of distinct lines or bands occupying definite positions, that is, with definite wave lengths. The line spectra are produced by atoms, and so are often referred to as **atomic spectra**, whereas the so-called band spectra, which can in fact often be resolved into large number of closely spaced lines (p. 563), are obtained from molecules. It is from atomic spectra in particular that information concerning the extranuclear electrons has been obtained.

Excitation of Spectra.—The production, or "excitation," of atomic spectra may be brought about in various ways: a compound of the element under investigation, e.g., a chloride of a metal, may be heated in a Bunsen flame when a characteristic color is frequently obtained, and on examining the flame in a spectroscope the line spectrum of the element is seen. If the element forms a compound which burns under suitable conditions with the production of a reasonably hot flame, then the atomic spectrum can often be observed. Spectra may also be excited by the passage of an arc through a gas or vapor, or by impregnation of the electrodes, between which the arc is struck, with a compound of the element. The spectrum of the material composing the electrodes can also be obtained in this manner. Flame and arc spectra are generally grouped together as being most easily excited, and the lines obtained are mainly due to normal atoms. By means of sparks, or by the passage of an electrical discharge, when relatively high potentials have to be employed, the spark spectra may be excited, in addition to the arc spectra. The discharge tube method of excitation is particularly useful for elements which are gaseous or fairly volatile at ordinary temperatures; the energy of the discharge is also generally sufficient to bring about dissociation of molecules so that the spectra of their constituent atoms can be observed. With increasing potentials new series of lines are frequently produced, the wave lengths being successively shorter, that is, further toward the ultraviolet region of the optical spectrum; it is, therefore, convenient to consider the first spark spectrum, the second spark spectrum, and so on.

The new spectra appearing when this method of excitation is used are

due to ionized atoms, that is, atoms which have lost one or more electrons under the influence of the powerful electric discharge, as distinct from the arc spectra produced by normal atoms. Evidence for this view is obtained from the resemblances observed between the detailed structures of the normal spectrum of sodium, the first spark spectrum of magnesium, and the second spark spectrum of aluminum. These three elements which succeed one another in the periodic table have atomic numbers of 11, 12 and 13, respectively, and consequently they have these numbers of electrons. If a magnesium atom loses an electron, becoming a Mg^+ ion, and an atom of aluminum loses two electrons, becoming an Al^{++} ion, the resulting ions have the same number of electrons as the normal sodium atom, and the similarity in the spectra is not surprising. Many cases of analogous resemblances are known, and the first and second spark spectra are generally attributed to singly and doubly ionized atoms, respectively; spark spectra of even higher orders have been observed, and these are produced by more highly charged ions. The results of an examination of arc and spark spectra have been generalized in the **spectroscopic displacement law** of W. Kossel and A. Sommerfeld (1919); it may be put in the following form. *The arc (normal atom) spectrum of an element is similar, especially as regards fine structure, to the first spark (singly ionized) spectrum of the element one place higher in the periodic table, to the second spark (doubly ionized) spectrum of the element two places higher, and so on.* The nomenclature frequently employed is to use Roman numerals, viz., I, II, III, IV, etc., for the arc, first spark, second spark, third spark spectra, etc., respectively; thus, Mg I, Mg II, and Mg III would represent, for spectroscopic designation, the normal magnesium atom, and Mg^+ and Mg^{++} ions, respectively.

The spectra of normal atoms can often be excited by bombardment of the vapor by electrons accelerated at relatively low potentials; this method is important as it permits of the gradual development of the spectrum under controlled conditions. For example, the use of electrons accelerated by a potential of 3.2 volts excites only one line of the magnesium spectrum; at 6.5 volts an additional line is present, but at 10 volts the normal arc spectrum of Mg I is obtained. By increasing the potential to 15 volts the electrons acquire sufficient energy to expel one of the extranuclear electrons of the magnesium atom, and the Mg II spectrum (first spark), that is, of Mg^+ ions, appears.

Emission and Absorption Spectra.—The spectra so far considered are of the emission type, lines of definite color being observed in the visible spectrum, and photographically active lines are obtained in the ultraviolet region. Many of these lines can also be studied in "absorption spectra," of which the well known Fraunhofer lines of the sun's spectrum are an example. If continuous radiation, such as white light, is passed through the vapor of an element, black lines, showing that light of the corresponding wave length has been absorbed, can be observed spectroscopically. The positions of these absorption lines correspond exactly with those of the visible lines of the emission spectrum of the

same element. Frequently more lines are observed in emission than in absorption, and so the latter type are often used where a simplification is desirable.

Spectral Series.—When the spectrum of a given element has been recorded on a photographic plate, the next problem is to sort out the lines into definite series. This is frequently a matter of great difficulty, as will be realized from the fact that on the plate there may appear several series of lines produced by the normal atom, and also those due to various ionized forms, as well as possibly bands resulting from molecules present. In some cases the spectra are so complex that complete analysis of the lines has not yet been achieved, but for elements of low atomic weight, the sorting process, although tedious, has generally been carried out successfully. The following devices are often used. First, all lines belonging to a given series are similar in appearance; for example, they may all have clearly defined edges or may all be diffuse. Second, all lines in the same series are resolved in a similar manner in a strong magnetic field (Zeeman effect, p. 63). Third, the distances between successive lines in a series decrease according to a more or less definite law, each series approaching a limit in the shorter wave length region.

The Spectrum of Hydrogen.—As might be anticipated the atomic spectrum of hydrogen is the simplest of line spectra. The well known H_α , H_β , H_γ , H_δ , etc., lines, which appear in the visible region, were first studied in the spectrum of the sun, and their wave lengths have been measured with great accuracy. In 1885, J. J. Balmer pointed out that the results could be expressed by means of the relationship

$$\lambda = k \frac{n^2}{n^2 - 4}, \quad (45)$$

where k is a constant, and n is an integer, equal to 3, 4, 5, 6, etc., for the H_α , H_β , H_γ , H_δ , etc. lines, respectively. This equation, which has become of fundamental importance, can be written in the form

$$\frac{1}{\lambda} = \bar{\nu} = R_H \left(\frac{1}{2^2} - \frac{1}{n^2} \right), \quad (46)$$

where $\bar{\nu}$, in cm^{-1} units, the reciprocal of the wave length, represents the number of wave lengths per $\text{cm}.$, and is generally referred to as the **wave number**.* The constant R_H is called the Rydberg constant, and is equal to $109,677.76 \text{ cm}^{-1}$. The remarkable exactness with which this simple formula reproduces the actual wave lengths, or wave numbers, of the Balmer series, as the hydrogen lines under discussion have been named, is shown by the data in Table 2, where the calculated and observed

* In spectroscopic work it is the common practice to express frequencies in terms of wave numbers, that is, the number of vibrations per $\text{cm}.$, $\bar{\nu} \text{ cm}^{-1}$, instead of in terms of the true frequency which is the number of vibrations per sec., $\nu \text{ sec}^{-1}$. The two quantities are related by the expression $\nu = \bar{\nu}c$, when c is the velocity of light.

wave lengths for various values of n are compared. It will be noted that as n increases the lines become closer together; the frequencies gradually approach a **convergence limit** which can be calculated by putting n equal to infinity in (46). The limit is thus at a wave number of $R_H/4$, that is 27,419.45 cm.⁻¹, or at a wave length of 3647.05 Å.

TABLE 2. COMPARISON OF OBSERVED AND CALCULATED WAVE LENGTHS OF BALMER LINES IN HYDROGEN SPECTRUM

n	Wave Length	
	Observed	Calculated
3	6562.79 Å.	6562.80 Å.
4	4861.33	4861.38
5	4340.47	4340.51
6	4101.74	4101.78
7	3970.06	3970.11
8	3889.00	3889.09
9	3835.38	3835.43

The Ritz Combination Principle.—An examination of the Balmer formula (46) shows that it consists of two terms: one of these is a constant, or **series term**, and the other is variable and is referred to as the **current term**. In 1908 W. Ritz introduced his **combination principle**, which is virtually a generalization of the Balmer formula. In its simplest form the principle states that *the wave number of any spectral line may be represented as the combination of two terms, one of which is constant and the other variable throughout each spectral series; thus,*

$$\bar{\nu} = \frac{R}{x^2} - \frac{R}{y^2}, \quad (47)$$

x and y being integral for hydrogen, but not necessarily for other atoms (cf. p. 40). The word **term** as used in spectroscopic work has the specific meaning implied in the Ritz combination principle, and is expressed in wave numbers; thus in (47), R/x^2 and R/y^2 are the two combination terms.

The combination principle applies throughout the whole range of spectroscopic studies, from X-rays to infrared, and has proved to be of fundamental importance, since it has led to the development of the important concept of energy levels. For the spectra of atomic hydrogen the frequency of the lines may be represented by the general equation

$$\bar{\nu} = R_H \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right), \quad (48)$$

where R_H is the Rydberg number, as given above, and n_1 and n_2 are integral, the former being constant throughout any given series. It has been seen that for the Balmer series n_1 is equal to 2, and as a result of the Ritz principle, F. Paschen (1908) sought for the series of lines for which n_1 had the constant value of 3, and found them in the infrared region of the spectrum; this series was later extended by F. S. Brackett (1922) and

A. H. Poetker (1927). In the ultraviolet spectrum of hydrogen, T. Lyman (1906, 1916) had observed a number of lines whose frequencies were also in excellent agreement with the requirements of (48), n_1 having the value of 1, with n_2 equal to 2, 3, 4, etc. In more recent years the series for which n_1 is 4 was discovered in the infrared by F. S. Brackett (1922), and the beginning of another series, with n_1 equal to 5, has been found by A. H. Pfund (1924) in the same region of the spectrum. The results may be summarized in Table 3.

TABLE 3. THE ATOMIC SPECTRUM OF HYDROGEN

Series	n_1	n_2
Lyman	1	2, 3, 4, ...
Balmer	2	3, 4, 5, ...
Paschen	3	4, 5, 6, ...
Brackett	4	5, 6, 7, ...
Pfund	5	6.

The remarkable fact emerges from these results that the frequency of any line in the hydrogen spectrum may be expressed by the combination of two terms, each of which may be represented by R_H/n^2 , where n is integral. The interpretation of this regularity was given in the first place by the theory of N. Bohr (1913), and later by means of wave mechanics; both these subjects, and also the problem of the fine structure of the hydrogen lines, will be considered shortly.

The Spectrum of Ionized Helium.—At one time two series of spectral lines, the frequencies of which could be expressed by the equations

$$\bar{\nu} = R \left(\frac{1}{1.5^2} - \frac{1}{m^2} \right) \quad \text{and} \quad \bar{\nu} = R \left(\frac{1}{2^2} - \frac{1}{(m + \frac{1}{2})^2} \right), \quad (49)$$

were attributed to hydrogen, but when the Bohr theory had provided a theoretical basis for the Ritz principle it was realized that this could not be the case. The lines under consideration have now been shown to be produced by ionized helium He^+ , which having only one extranuclear electron is similar to the hydrogen atom. The equations (49) are now rewritten as

$$\bar{\nu} = 4R_{\text{He}} \left(\frac{1}{3^2} - \frac{1}{n^2} \right) \quad \text{and} \quad \bar{\nu} = 4R_{\text{He}} \left(\frac{1}{4^2} - \frac{1}{n^2} \right), \quad (49a)$$

where R_{He} is 109,722.40 cm^{-1} , and so is slightly larger than the Rydberg number for hydrogen. Two other series, in which the first term in the parentheses has the values $1/1^2$ and $1/2^2$, respectively, were observed by T. Lyman (1924) and by K. T. Compton (1928), so that the spectrum of singly-ionized helium is quite analogous to that of atomic hydrogen. It may be pointed out that the number 4 preceding the Rydberg constant in (49a) is the square of the **effective charge**, which is 1 for a neutral atom, 2 for a singly charged ion, 3 for a doubly charged ion, and so on.

Doubly-ionized lithium, Li^{++} , and trebly-ionized beryllium, Be^{+++} , both consist of a single electron and a positively charged nucleus; they resemble hydrogen and singly-ionized helium in structure, and are said to be **hydrogen-like** or **hydrogenic** in character. The frequencies of a number of lines in the spectra of Li^{++} and Be^{+++} have been measured and found to comply with equations analogous to those for hydrogen and He^+ , except that the Rydberg number is preceded by 9 and 16, respectively, these being the squares of the respective effective charges, viz., 3 and 4. In general the frequencies of the lines in the spectra of hydrogen-like atoms may be represented by the formula

$$\bar{\nu} = Rz^2 \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right), \quad (50)$$

where z is the effective charge of the atom, or ion, as defined above, and R is the appropriate Rydberg constant which varies slightly from one atom to another.

Spectra of Higher Elements.—As a result of the examination of atomic spectra it has been concluded that for many atoms, other than the hydrogenic type, especially if they belong to the first three groups of the periodic table, or are brought by ionization into these groups, the lines can be divided into at least four distinct series. These are referred to as the **Sharp, Principal, Diffuse** and **Fundamental** series, respectively. Some of the names are not satisfactory, since, for example, the lines in the Sharp and Diffuse series are not always as described by the particular adjective employed, but the chief discrepancy lies in the use of the word “fundamental”; the series is not really fundamental and so is often referred to as the Bergmann series. The names given above will, however, be retained because the initial letters, viz., *S*, *P*, *D* and *F*, have been employed to designate the series, as well as for other purposes. In each of the series mentioned the frequencies of the lines may be represented by a combination of a series (constant) and a current (variable) term, in accordance with the Ritz principle. The frequency, i.e., wave number, equation can be written

$$\bar{\nu} = L - \frac{Rz^2}{(n - \delta)^2}, \quad (51)^*$$

where R is the Rydberg number, which attains a constant value of $109,737 \text{ cm.}^{-1}$ with increasing atomic weight of the element, and z is the effective charge, as already described. For alkali-like atoms, viz., normal atoms of the first group, and for nonionized atoms of any group of the periodic classification, z is equal to unity. The integer n is modified by the

* Many writers, following Rydberg, use $n + \delta$, or equivalent symbols, instead of $n - \delta$ as given here; the latter representation is preferred as it suggests immediately that the quantity is less than the integer n . It may be noted that L can be expressed in the same form as the variable term involving, of course, different values of n and δ ; equation (51) would then be exactly analogous to the Ritz equation (47).

quantity δ which is approximately, although not exactly, constant for any given series; for alkali-like atoms, however, δ may be assumed constant. If n is put equal to infinity the variable term of (51) becomes zero, and hence $\bar{\nu}_\infty$ is equal to L , the constant term; the latter thus gives the frequency of the convergence limit of the particular spectral series.

Analysis of the lines of the Sharp series shows that the terms necessary to express their frequencies may be represented by the symbol nS , where the smallest value of n depends on the nature of the alkali metal, e.g., 3 for lithium, 4 for sodium, etc. In the Principal series the terms may be written as nP , the minimum value of n being 2 for lithium, 3 for sodium, etc. The lowest values of n in the terms nD for the Diffuse series for lithium and sodium are 3 in each case. Finally, in the Fundamental series the smallest n in the term nF for both lithium and sodium is 4.

Experimental study of the frequencies of the lines in the various spectra has shown that the constant terms of the Sharp and Diffuse series are in each case equal to the first current term of the Principal series. Further, the constant term of the Fundamental series is equal to the first current term of the Diffuse series, whereas that of the Principal series is equivalent to a nonexistent term, of the Sharp series, e.g., $2S$ for lithium, $3S$ for sodium, etc. The frequencies of the lines of the four series may consequently be represented by the following general expressions:

Sharp series	$\bar{\nu} = aP - nS$
Principal series	$\bar{\nu} = aS - nP$
Diffuse series	$\bar{\nu} = aP - nD$
Fundamental series	$\bar{\nu} = aD - nF,$

where the lowest values of the integers in the constant terms (a) and current terms (n) in each case have been indicated above.

In addition to the lines of the four series described, others are frequently observed in atomic spectra; these sometimes arise from other combinations of the terms given above, or from further series of terms, to which the symbols G , H , etc., have been attached. The actual combination of terms is, however, limited by a definite selection principle which will be considered in connection with the theories of atomic spectra (p. 60). There are also other complexities which need not be discussed here as they have apparently no direct physicochemical application.

Fine Structure of Spectral Lines.—It has been tacitly assumed hitherto that all the lines in the spectral series under discussion are single, but closer examination shows that many of them have what is called a **multiplet structure**, that is to say, the spectrum consists of groups of two, three, or more, lines so very close together that unless a spectroscope of high resolving power is used each group appears as a single line. A simple illustration of this behavior is provided by the familiar yellow

D-line of sodium, which is the first in the Principal series of this element; careful examination has shown that it is actually a closely spaced doublet, that is, it consists of two separate lines. In general, the spectral lines for all the alkali metals are doublets, the frequency separation of the two lines being almost constant in the Sharp and Diffuse series, while in the Principal series the separations become gradually smaller and the two lines coincide at the convergence limit. With the alkaline earth metals, viz., the second group of the periodic table, many of the lines are closely spaced groups of three, i.e., triplets, whereas others are singlets. The lines of the elements in the third group of the periodic classification are doublets and, in some cases, quartets, whereas the elements in the fourth group show multiplicities of one, three and even five. These relationships apply, of course, to the arc spectra, that is to the spectra of the normal atoms; in its first spark spectrum, however, as already noted, an element behaves as if it were in the next lower group of the periodic table, and so on. The problem of spectral multiplicity, its theoretical basis, and the law relating multiplicity to the position in the periodic table will be considered on p. 66.

Origin of Spectra.—When the atomic volumes of the elements in the solid state are plotted against their atomic weights, or atomic numbers, there is obtained the familiar Lothar Meyer curve with its series of waves corresponding to the periods of the periodic table. This result suggests that many physical properties, which recur periodically among the elements, are related in some manner to the volume occupied by the atom. According to the nuclear theory, it is to be expected that this would be determined to a great extent by the arrangement of the electrons around the nucleus, and so it is probable that the electronic structure of the atom is responsible for many periodic physical properties. It has been seen that the multiplicities of the lines in atomic spectra correspond to the groups of the periodic table, and, in general, elements in the same group have similar spectra; in fact, analysis of the atomic spectrum of one element in a group can be used as a basis for the analysis of another in the same group. The spectrum of an element may, therefore, be regarded as a periodic property, and it is easy to understand, therefore, the development of the view that atomic spectra are related in some way to the grouping of electrons in the atom.

Since the physical properties which are related to the electronic arrangement are periodic in nature, it is probable that certain groupings of electrons recur periodically. In proceeding through the series of elements, the number of electrons increases by unity from one atom to the next in the order of increasing atomic number, and it appears, therefore, since definite arrangements recur from time to time, that the electrons arrange themselves in groups or shells, with the same number in excess of completed shells at each corresponding element. For example, the alkali metals probably all have one electron in excess, the alkaline earth metals

have two electrons in excess of completed groups, and so on. There is reason to believe, as will be seen subsequently, that these extra electrons determine many chemical properties, and in particular the valence, of an element, as well as the main characteristics of the atomic spectrum, and so they are often called the **optical electrons** or **valence electrons** of the atom. On the basis of these arguments it has been possible to develop important theories concerning the origin of spectra, and use has been made of the observed spectra of individual elements to suggest possible electronic arrangements.

Bohr's Theory of Atomic Spectra.—To account for the fact that the electrons did not fall into the nucleus as a result of electrostatic attraction, E. Rutherford (1911) found it necessary to postulate rapid rotation about the nucleus, so that the attractive force inward was exactly balanced by the outward centrifugal force. It was pointed out, however, that such a system cannot be stable. The analogy between a planet rotating round the sun and an electron rotating about a nucleus is fallacious, because the electron and nucleus are charged bodies. Any particle moving with a constant speed v in a circle of radius r possesses an acceleration v^2/r towards the center, but if the particle is an electron then classical electromagnetic theory requires that it should emit radiation during its motion, since the energy of the electromagnetic field associated with an accelerated electron is continually changing. According to the simple theory, therefore, the electron will be radiating energy continuously, and the radius of curvature of its path would steadily decrease; the electron should then follow a spiral path and eventually fall into the nucleus. There is, further, the difficulty of explaining the emission of definite spectral lines which, according to the quantum theory (p. 32), should correspond to definite amounts of energy. The spectrum of the atom would be related to the energy radiated by the moving electron, but this energy would be changing as the radius of curvature of its path, and hence its acceleration, changed and so the spectrum might be expected to be a continuous one covering all frequencies, instead of consisting of lines falling into well-defined series.

In order to overcome these difficulties N. Bohr (1913) made the revolutionary suggestion that an electron always moved in a closed orbit, and as long as it did so there was no absorption or emission of radiation. In other words, contrary to the requirements of classical theory, it was postulated that an electron can have acceleration and yet not radiate energy. More than one stable orbit was considered to be possible for any electron, each being termed a **stationary state**; at the present time these are spoken of as the **energy levels** of the electron or of the atom. According to Bohr's theory, the number of such states or energy levels is determined by the quantum condition that the angular momentum of the electron is an integral multiple of $h/2\pi$, where h is the Planck constant (p. 32). At the time this view was proposed, there was no reason for

the "quantization" of momentum in terms of $h/2\pi$ units, except that the correct results were obtained by employing this postulate. Wave mechanics, however, as will be seen later, provides the necessary justification. The angular momentum of a particle moving in a circular orbit is mvr , where m is its mass, and v and r have the same significance as above; hence it follows, from Bohr's quantum postulate, that the possible orbits must be defined by the relationship

$$mvr = n \frac{h}{2\pi}, \quad (52)$$

where n is an integer, called the **quantum number** of the particular energy level. It will be seen from (52) that as n increases so does r , the apparent radius of the electronic orbit; the quantum numbers may thus be regarded as giving the order of the various orbits, counting from the nucleus outwards. For the first possible orbit, or stationary state, n would be equal to 1, for the second n would be 2, and so on.

The final postulate of Bohr's theory is that when an electron passes suddenly, or "jumps," from one orbit to another, radiation of a definite frequency, giving a definite spectral line, is emitted or absorbed. The frequency is dependent on the energies of the initial and final states in the following manner. If E' is the energy of the electron in the orbit from which it starts, and E'' is the value in the orbit to which it jumps, then Bohr suggested that the frequency (ν sec.⁻¹) of the spectral line produced is related to the energy change $E' - E''$ by the quantum relationship

$$E' - E'' = h\nu, \quad (53)$$

or, expressing the frequency in terms of wave numbers ($\bar{\nu}$ cm.⁻¹),

$$E' - E'' = hc\bar{\nu}; \quad (54)$$

$$\therefore \bar{\nu} = \frac{E' - E''}{hc}, \quad (55)$$

since ν is equal to $\bar{\nu}c$ (p. 37). Each particular energy transition, from one orbit to another, should thus result in the formation of a definite spectra line whose frequency is given by (55).

In general, if atoms of a substance are exposed to conditions under which energy absorption, that is **excitation**, is possible, for example, by the use of high temperature or an electrical discharge, the electrons will take up energy, and various jumps from lower to upper energy levels, that is, from inner to outer orbits, will occur. The return of the electrons from the various upper levels to the lower levels will result in the liberation of definite amounts of energy, and each transition corresponds to a line of definite frequency in the emission spectrum. Since a number of different energy levels are possible in every atom, there will be a number of transitions and hence a series of lines will be observed in the spectrum. The

production of absorption spectra can be accounted for in an analogous manner: the various amounts of energy corresponding to the frequencies of the absorption lines are those required to raise the electrons from lower to higher levels, each line corresponding to a definite transition.

Comparison of (55) with (17) shows that Bohr's use of the quantum theory provides an interpretation of the Ritz equation. The two terms of the latter are evidently related to the energy of the atom in the initial and final states involved in an electronic transition. When one term is constant throughout a given spectral series, while the other varies, as in (51), it means that in this particular series all the transitions are to (or from) one specific energy state. Spectral terms can thus be identified with electronic energy levels (or states), and hence the notation used to describe a particular term, e.g., $2S$, $3P$, etc., is also applied to the atom with the optical electron (or electrons) in the corresponding energy state. In the P state of sodium, for example, the optical electron is in such an energy level that every transition to (or from) this level would contribute a P term to the Ritz equation. Various P states are possible, these being identified by the value of n (p. 41). The significance of n and the term letter will be explained later.

Limitations of the Bohr Theory.—So far the Bohr theory has been applied here only to account for the formation of the line spectra of atoms, but by consideration of the quantitative aspect of the postulates it is possible to calculate with remarkable accuracy some of the main features of the spectrum of hydrogen. Before proceeding to consider these, it is desirable to call attention to the fact that according to modern quantum mechanics the Bohr model of electrons rotating in definite *orbits* cannot be correct, although there is little doubt that every atom must have a definite set of electronic energy *levels* or *states*. Unfortunately, the modern treatment has to regard the atom as a purely mathematical concept without requiring any definite model as a basis for the calculations. The Bohr theory is essentially pictorial, giving physical significance to the electron as a particle moving in one of a series of possible orbits, but what it gains in pictorial value it loses in exactness; this dilemma is inevitable when the interior of the atom is under consideration. In the treatment above it has been assumed that both the momentum and position, in a definite orbit, of an electron are known, but the uncertainty principle (p. 18) shows this to be impossible. If the electron momentum is known exactly, then there must be an uncertainty with regard to the position of the orbit, and the extent of this indefiniteness is appreciable in comparison with the size of the atom. In spite of this fundamental limitation, the Bohr model provides a useful basis for elementary calculations, and it will be many years before it is discarded for this purpose. It is, however, important to bear in mind in the discussion which follows that an attempt is being made to give reality to a state of affairs within the atom to which as yet, if ever, no reality can be given.

Hydrogen-like Atoms.—The simplest case to consider is that of the so-called hydrogen-like atoms (p. 40) consisting of only one electron rotating in a *circular* orbit about the nucleus which has an effective charge of z units, that is to say, the virtual charge is ze , where e represents the

unit of charge, namely that of the electron. The total energy E of the electron is the sum of its kinetic and potential energies; the former of these is $\frac{1}{2}mv^2$, as for an uncharged body, where m is the mass of the electron and v its velocity. If the potential energy of an electron is taken as zero when it is at an infinite distance from the nucleus, the value at a distance r is given by $-ze^2/r$, the negative sign indicating that work must be done *on* the electron in order to remove it to infinity.* The total energy is, therefore, given by

$$E = \frac{mv^2}{2} - \frac{ze^2}{r}. \quad (56)$$

As long as the electron moves in a stationary orbit, the centrifugal force outward must be equal and opposite to the electrical attraction inward; it has been noted (p. 43) that the acceleration of the electron toward the center is v^2/r , so that the centrifugal force must be mv^2/r . Assuming the Coulomb law of inverse squares to apply, the electrical attraction between the nucleus and the electron is ze^2/r^2 , so that

$$\frac{mv^2}{r} = \frac{ze^2}{r^2}, \quad (57)$$

and, therefore,

$$\frac{mv^2}{2} = \frac{ze^2}{2r}. \quad (58)$$

By substituting for $mv^2/2$ in (56) it is seen that

$$E = -\frac{ze^2}{2r} = -\frac{mv^2}{2}, \quad (59)$$

the negative sign implying, as before, that energy must be supplied to the electron to move it to an infinite distance from the nucleus.

The condition that the angular momentum is quantized in units of $h/2\pi$ is now applied by means of (52), from which there may be derived

$$\frac{mv^2}{2} = \frac{nhv}{4\pi r}. \quad (60)$$

By comparison with (59) it follows that

$$\frac{ze^2}{2r} = \frac{nhv}{4\pi r}, \quad (61)$$

* By the inverse square law the attractive force on the electron is $ze \times e/r^2$; if it is moved through a small distance dr the work done is $(ze^2/r^2)dr$. To obtain the potential energy at a distance r from the nucleus integrate between the limits r and infinity, noting that when r is infinite the potential energy of the electron is zero; the result is $-ze^2/r$, as given above.

and hence

$$v = \frac{2\pi ze^2}{hn}. \quad (62)$$

If this value of v is inserted in (59), the result

$$E_n = - \frac{2\pi^2 z^2 e^4 m}{h^2 n^2} \quad (63)$$

is obtained, where E_n represents the energy of the electron in the circular orbit characterized by the quantum number n , that is, when its angular momentum is $nh/2\pi$. When an electron jumps from an outer orbit, in which its quantum number is n_2 , to an inner orbit in which it is n_1 , the energy emitted as radiation is given by

$$E_{n_2} - E_{n_1} = \frac{2\pi^2 z^2 e^4 m}{h^2} \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right) \quad (64)$$

and by (55), the frequency, expressed in wave numbers, of the corresponding line in the emission spectrum is obtained on dividing by hc ; thus,

$$\bar{\nu} = \frac{2\pi^2 z^2 e^4 m}{h^4 c} \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right) \quad (65)$$

$$= R z^2 \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right), \quad (66)$$

where R is a universal constant, since it involves numerical constants only, and is in fact equal to the Rydberg number for atoms of high atomic weight (p. 48).

Actually (66) is not quite complete, for the motion of the nucleus has not been taken into account. Whenever a particle moves in a circle about another of finite mass, then according to the laws of dynamics, the central particle must also move in a circle about the common center of gravity. In order to make allowance for the relative motion of the nucleus the mass m of the electron must be replaced by the "reduced mass" of the atom, viz., $mM/(m + M)$, where M is the mass of the nucleus; the right-hand side of (66) should thus be multiplied by $M/(m + M)$, so that

$$\bar{\nu} = \frac{M}{m + M} R z^2 \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right). \quad (67)$$

This equation is clearly of the correct form, as required by the Ritz combination principle, to represent the spectra of hydrogen-like atoms with a single optical electron, for example, atomic hydrogen, singly ionized helium He^+ , doubly ionized lithium Li^{++} , and trebly ionized beryllium Be^{+++} . By comparing (67) and (50) it is seen that from the Bohr theory it is possible to deduce a relationship for the frequency of

spectral lines involving the square of the effective charge, as actually found in practice for atoms and ions of the type under consideration. The small variations in the Rydberg number are explained by the factor $M/(m + M)$, but as this approaches unity with increasing weight of the nucleus, it is evident that a constant value, equal to R , will be approached as the atomic weight increases.

A quantitative test of (67) may be made by inserting the known values for the various quantities involved in the factor $Rz^2M/(m + M)$, and seeing how closely it approaches the known value for the Rydberg number to which it corresponds. For hydrogen, for example, z is 1, and the mass M of the nucleus is 1.673×10^{-24} g.; the other quantities are universal constants, viz., e is 4.802×10^{-10} e.s. unit, m is 9.107×10^{-28} g., h is 6.624×10^{-27} erg sec., and c is 2.9977×10^{10} cm. per sec.; hence,

$$R_H = \frac{2\pi^2 e^4 m}{h^3 c} \cdot \frac{M}{m + M} = 1096 \times 10^2 \text{ cm.}^{-1},$$

which, considering the limited accuracy of the data for the various constants, is in remarkably good agreement with the experimental value, 109,677.8 cm.⁻¹. For heavier atoms the Rydberg constant should be somewhat greater, approaching a limiting value as the atomic weight increases. Actually the Rydberg numbers for hydrogen and singly ionized helium are known with much greater accuracy than are the constants used in the calculation given above, so that the experimental numbers (pp. 37, 39) may be used to determine atomic constants. The ratio of the Rydberg numbers for hydrogen and singly ionized helium may be written, according to (67), as

$$\frac{R_H}{R_{He}} = \frac{109,677.76}{109,722.40} = \frac{(m + M_{He})M_H}{(m + M_H)M_{He}}, \quad (68)$$

where M_H and M_{He} are the masses of the respective nuclei. If it is assumed that the helium nucleus has approximately four times the mass of the hydrogen nucleus, it follows that

$$\frac{M_H}{m} \approx \frac{\frac{1}{4}R_{He} - R_H}{R_H - R_{He}} = 1839. \quad (69)$$

The mass of the proton is thus about 1839 times that of an electron, so that the mass of an electron is 1/1840th part of that of a hydrogen atom, in agreement with the value 1/1838 given on p. 6. Using the latter ratio and the known Rydberg constant for hydrogen, 109,677.76 cm.⁻¹, the accurate limiting value of R , the Rydberg number for high atomic weights, may be calculated as 109,737.4 cm.⁻¹.

It is seen that the Bohr theory is able to account in a remarkable manner for the exact positions of the main lines in the spectra of hydrogen-like atoms or ions, e.g., He⁺, Li⁺⁺ and Be⁺⁺⁺. Further, it gives an ex-

planation of the fundamental basis of the Ritz combination principle that the frequency of any spectral line can be represented as the difference of two terms. According to the theory, the term values for a given line when multiplied by hc give the (negative) energies of the two electronic states responsible for that line, as may be seen from a consideration of (64), (65) and (66). It should be noted that since the innermost levels have the highest (negative) energies, they will also correspond to the highest term values.¹⁵

The Hydrogen Spectrum.—For all the lines in the Lyman series n_1 in (67) is equal to 1 (cf. p. 39); hence, according to the Bohr theory, these

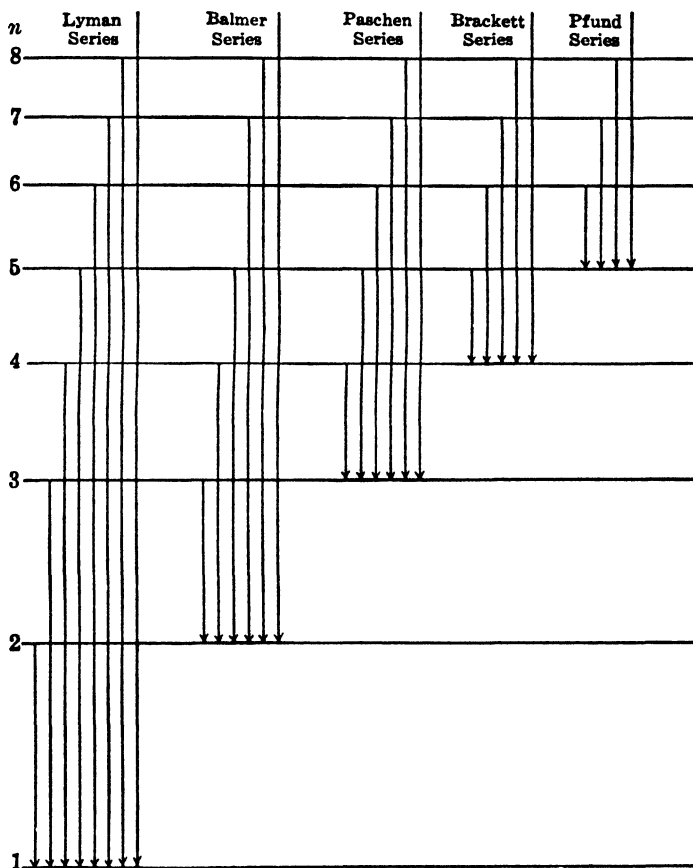


FIG. 10. Formation of emission spectra of atomic hydrogen

lines, in the emission spectrum, are produced when an electron jumps from *any* higher level to the lowest level, or innermost orbit, of the hydrogen atom, for which n is 1. When excited, for example by an electric dis-

charge, the electrons in different atoms will not all be raised to the same level; they will, in fact, be distributed over a number of levels with n equal to 2, 3, 4, 5, 6, etc. Consequently, when the electrons fall back to the lowest level, where n_1 is 1, there will be produced a *series* of spectral lines, and not merely one line, corresponding to n_2 values in (67) of 2, 3, 4, 5, 6, etc., thus giving the Lyman series. Similarly, the electrons in various hydrogen atoms falling back, after excitation, from levels of quantum numbers (n_2) equal to 3, 4, 5, 6, etc., to that in which it (n_1) is 2, will produce the lines of the Balmer spectrum (cf. p. 37). The general basis of the formation of the spectrum of atomic hydrogen may be depicted diagrammatically in Fig. 10; the arrows represent the return of electrons from upper to lower quantum energy levels, and so indicate the method of formation of the various emission spectra. The production of the corresponding absorption spectra may be represented by the same transitions in the opposite direction.

Critical Potentials.—Confirmation of the concept of stationary energy levels and of the quantization of energy in the atom has been obtained from the study of **critical potentials**. In general, a critical potential is a measure of the amount of energy required to raise an electron in an atom from a lower to a higher level. There are two distinct kinds of critical potentials, known as **ionization potentials** and **resonance potentials**. The ionization potential represents the work necessary to remove an electron from a normal atom, wherein the electron may be supposed to be in its lowest level, to an infinite distance, so that a positively charged ion results. The term "potential" is used because the energy quantity is measured by means of electrons accelerated by application of a known potential: the energy of the electrons is given by the product of the electronic charge and the accelerating potential, and so is expressed in electron-volts, frequently abbreviated to "volts." The resonance potential, which is determined in a similar manner and also quoted in volts, is a measure of the work required to raise an electron from its lowest level to any other level. A series of potentials, known as the first, second, etc., resonance potential, can be obtained corresponding to the transfer of an electron from the lowest to the next, next-but-one, etc., level. After excitation to the first resonance level, the electron on returning to its normal state emits radiation consisting of a single frequency, apart from fine structure; this is known as the first resonance line of the element. The second, third, etc., resonance lines are defined in an analogous manner. The familiar yellow line of sodium, wave length 5890 Å., and the important line in the ultraviolet region emitted by mercury, at 2537 Å., are both examples of first resonance lines.

Measurement of Critical Potentials.—Several methods of direct measurement have been devised, but the most common involve the same principle. If an electron of low energy collides with an atom, the collision is generally **elastic**; that is to say, only translational, and not internal, energy is exchanged between

the particles. As the energy of the electron is increased, by raising the accelerating potential, a point is reached when it is sufficient to raise the energy of the atom with which it collides, from a lower to a higher energy level. There is then the possibility of an **inelastic collision**, the kinetic energy of the bombarding electron being converted into internal energy of the atom. If this occurs the electron will be deprived of nearly all its energy, and its speed, which has been steadily increasing as the accelerating voltage is increased, will suddenly fall to a small value. The voltage at which this change occurs will be equal to one of the critical potentials of the atom.

The method used by J. Franck and G. Hertz (1914) gave the first reliable data on critical potentials; the apparatus, shown diagrammatically in Fig. 11, consists of three electrodes, somewhat similar to those in a vacuum tube. The filament F , heated by an electric current, acts as a source of electrons which are accelerated towards the plate P , in the direction shown by the arrow, by means of a variable potential V_1 ; a small constant retarding potential V_2 , about 0.5 volt, acting in the opposite direction, is applied between the plate and the perforated screen S . The arrangement of electrodes and of gas pressure is such that an electron moving from F to S will probably make many collisions with atoms of the vapor present in the apparatus. As soon as the accelerating potential V_1 exceeds V_2 , electrons will commence to pass through the screen, and a flow of current will be indicated on the galvanometer G ;

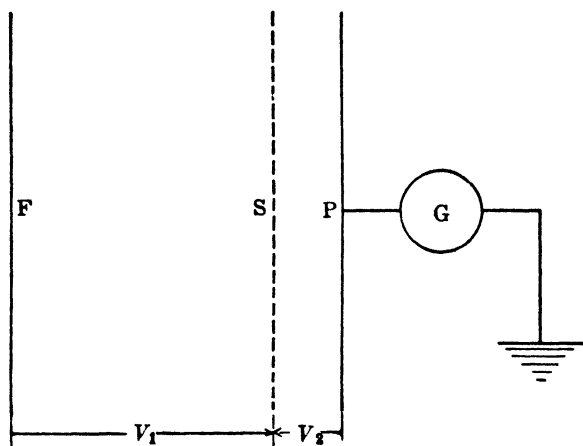


FIG. 11. Determination of critical potentials (Franck and Hertz)

the current strength will increase as V_1 increases, provided the electrons undergo elastic collisions only. When V_1 becomes equal to a critical potential, however, the electrons reaching S will just have attained sufficient energy to be involved in inelastic collisions. If this occurs their kinetic energy is reduced to zero and the retarding potential V_2 will be sufficient to turn them back. The current as indicated on the galvanometer will consequently undergo a sudden fall; the value of the potential V_1 at this point is equal to a critical potential V_c of the atoms present. When the accelerating potential exceeds V_c the electrons acquire sufficient energy to suffer inelastic collision *before* reaching S , and hence if they are robbed of an amount of kinetic energy corresponding to the critical

value, sufficient will remain to overcome the retarding potential V_2 , and so the current will increase again. When the value of V_1 is equal to $2V_c$, however, a large number of electrons will undergo two inelastic collisions between F and S , the second being near S , so that the retarding potential will prevent their access to P , and there will be another fall of current. This will be followed by another increase and a further decrease at $3V_c$, and so on. With mercury vapor in the apparatus the falls in the current were observed at potentials of 4.9, 9.8, 14.7 volts, indicating a critical potential of 4.9 volts; more accurate work has shown that there are two such potentials at 4.9 and 6.7 volts, and all the breaks in the current-potential curve occur at multiples of these values or of their sums. The results obtained point definitely to the conclusion that the mercury atom is capable of taking up energy in integral multiples of 4.9 and 6.7 e.v. only, and so provide very strong evidence for the quantum theory.

By spectroscopic examination it has been found that the attainment of the first critical, i.e., resonance, potential generally corresponds, as is to be expected from the quantum theory of spectra, to the production of the first resonance lines of the spectra of the atoms present. For example, at 4.9 volts mercury vapor was observed to emit the line of wave length 2537 Å., whereas sodium has a resonance potential at 2.1 volts and the 5890 Å. line appears when sodium vapor is bombarded with electrons accelerated by this potential. In accordance with the requirements of the quantum theory the energies of the lines, as given by $h\nu$, are equal, within the limits of experimental error, to the energies of the bombarding electrons. For the first resonance line of sodium, for example,

$$h\nu = hc/\lambda = 6.62 \times 10^{-27} \times 3 \times 10^{10}/5890 \times 10^{-8} = 3.37 \times 10^{-12} \text{ erg},$$

whereas the energy equivalent of the resonance potential is given by

$$2.1 \times 1.60 \times 10^{-12} = 3.36 \times 10^{-12} \text{ erg},$$

since 1 e.v. is equal to 1.60×10^{-12} erg (p. 6). Equally good agreement is obtained for the first resonance potential of mercury: the energy corresponding to the 2537 Å. line is 7.83×10^{-12} erg, and the equivalent of 4.9 e.v. is 7.84×10^{-12} erg.

The experimental method of Franck and Hertz has been improved in many ways, but it suffers from the disadvantage of not distinguishing between resonance and ionization potentials. This difficulty has been overcome by B. Davis and F. S. Goucher (1917); the principle involved may be illustrated by Fig. 12. Two screens S_1 and S_2 are employed, and three potentials are applied which accelerate electrons in the directions shown; V_1 is variable and is gradually increased, but V_2 and V_3 are fixed; V_3 is very small whereas V_2 must be larger than V_1 . As long as V_1 is less than V_2 no current flows, for the electrons are turned back by V_2 . When it is equal to V_c resonance radiation is produced in the space between F and S_1 , and this falling on the plate P causes photoelectrons to be emitted, but they are returned by the potential V_3 and so contribute nothing to the current. Some of the light is, however, reflected on to the side of S_2 nearest to P , and the photoelectrons so produced are accelerated by V_3 and cause a small increase of current. Every resonance potential is thus indi-

cated by a further increase of current. When V_1 attains the ionization potential, *positive ions* are formed near S_1 and they are accelerated by V_2 towards P , and passing through S_2 reach P ; there is now a large increase of current in a

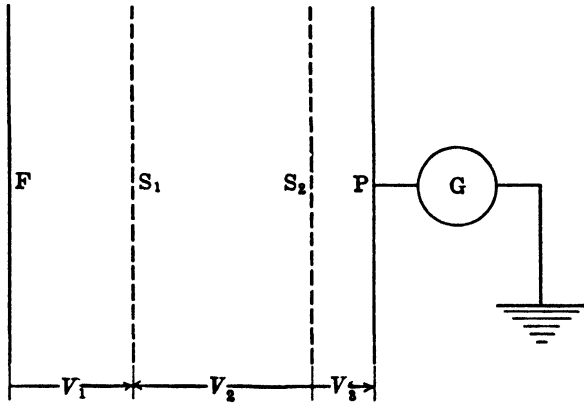


FIG. 12. Distinction between resonance and ionization potentials (Davis and Goucher)

direction *opposite* to that produced by the photoelectrons. With mercury vapor, for example, small breaks in current were observed at the resonance potentials of 4.9 and 6.7 volts, but there was a marked reversal at 10.4 volts, which is evidently the ionization potential

Calculation of Critical Potentials.—Any critical potential for a hydrogen-like atom may be calculated by inserting the correction term for relative motion of the nucleus in (64); thus,

$$W_{n_1, n_2} = E_{n_2} - E_{n_1} = \frac{2\pi^2 z^2 e^4 m}{h^2} \cdot \frac{M}{m + M} \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right) \quad (70)$$

$$= Rhcz^2 \frac{M}{m + M} \left(\frac{1}{n_1^2} - \frac{1}{n_2^2} \right), \quad (71)$$

where W_{n_1, n_2} represents the amount of work required to raise the electron from the level of quantum number n_1 to that in which it is n_2 , and R has its limiting value, viz., $109,737.4 \text{ cm.}^{-1}$. Since the ionization energy is the work required to remove an electron from its lowest level to infinity, the values of n_1 and n_2 are 1 and infinity, respectively; hence, for a hydrogen-like atom,

$$\text{Ionization energy} = Rhcz^2 \left(\frac{M}{m + M} \right). \quad (72)$$

For hydrogen itself, z is unity and the factor in parentheses is $1837/1838$; since R , h and c are known, the ionization energy may be evaluated as $2.178 \times 10^{-11} \text{ erg}$, which is equivalent to 13.59 e.v. From the Bohr treatment, therefore, the ionization potential of hydrogen should be 13.59 volts, and this is very close to the experimental value.

It was mentioned on p. 41 that when one of the term numbers, now known to represent a quantum number, was put equal to infinity, the frequency obtained was that of the convergence limit of the series. It will be evident, therefore, that if the frequency of the convergence line, in wave numbers, is multiplied by hc , the result will be the ionization energy, expressed in ergs. It is, of course, necessary to choose the convergence point in the correct spectral series, namely the one in which the lower electronic level is that corresponding to the state of the normal atom; for example, for hydrogen the convergence limit of the Lyman series must be taken, for it is only in this series that n_1 is equal to unity, the electronic level of hydrogen in its normal or **ground state**. The observed limiting frequency is $109,678 \text{ cm.}^{-1}$; multiplication by hc gives $2.178 \times 10^{-11} \text{ erg}$, so that the ionization potential is 13.59 volts, as quoted above. The method of estimating ionization potentials from the convergence point of the appropriate spectral series, either determined experimentally or from the correct form of (51), gives results in excellent agreement with those actually observed (p. 51). The series limits (1S) for the alkali metals, and the ionization potentials calculated from them, are recorded in Table 4, together with the observed values. Since ionization potentials are often difficult to determine experimentally, the values are frequently obtained from spectral data.

TABLE 4. OBSERVED AND CALCULATED IONIZATION POTENTIALS OF ALKALI METALS

Atom	Series Limit Cm. ⁻¹	Ionization Potential in Volts	
		Calc.	Obs.
Lithium	43,484	5.390	—
Sodium	41,449	5.138	5.13
Potassium	35,006	4.339	4.1
Rubidium	33,685	4.175	4.1
Cesium	31,407	3.893	3.9

The resonance potential of a hydrogenic atom can be evaluated from (71) by putting n_1 equal to 1, and n_2 to 2 for the first resonance potential, 3 for the second, and so on; thus,

$$\text{First resonance energy} = Rhcz^2 \frac{M}{m + M} \left(\frac{1}{1^2} - \frac{1}{2^2} \right) \quad (73)$$

$$= \frac{3}{4} Rhcz^2 \left(\frac{M}{m + M} \right), \quad (74)$$

so that the first resonance potential of a hydrogen-like atom is three-fourths of the ionization potential [cf. equation (72)]. For hydrogen itself this should be 10.19 volts, in agreement with the resonance potential obtained experimentally.

Like ionization potentials, resonance potentials can be calculated from spectroscopic frequencies; it has been seen that for mercury and sodium the energy equivalent of the first resonance potential is the same as that

of the resonance line, so that it is obviously possible to calculate the correct value of the potential from the frequency of the line. The lines in the Lyman series of hydrogen, for which n_1 is 1, are the resonance lines, and from these the resonance potentials of the hydrogen atom have been calculated; the results are in excellent agreement with those obtained by direct measurement.¹⁶

Radii of Hydrogen Orbits.—It is readily deduced from (52) that

$$v^2 = \frac{n^2 h^2}{4\pi^2 r^2 m^2}, \quad (75)$$

whereas from (57), since z is unity for hydrogen,

$$v^2 = \frac{e^2}{mr}, \quad (76)$$

and by combining (75) and (76), it is found that

$$r = \frac{n^2 h^2}{4\pi^2 m e^2}. \quad (77)$$

For the first Bohr orbit, that is when n is 1, it follows that

$$r = 0.53 \times 10^{-8} \text{ cm.},$$

and this should represent the radius of the normal hydrogen atom, the electron occupying the lowest level. In general, the radius of the n th orbit should be given by $0.53 \times 10^{-8} \times n^2$ cm., but it is doubtful, in view of the uncertainty as to the real physical meaning of the so-called electron orbits, if the results obtained in this way have any significance. The radius of the first orbit is of the same order as the radius of the hydrogen atom, determined from the spectrum of the molecule, and the modern wave-mechanical treatment shows that when n is unity the value of r in (77) probably gives a close approximation to the mean radius of the hydrogen atom.

Elliptic Orbits: Azimuthal Quantum Numbers.—The Bohr theory as outlined here gives the positions of the lines in the hydrogen spectrum, but it does not account for their fine structure as observed with spectroscopes of high resolving power. The theory was extended by A. Sommerfeld (1915) in the following manner. Hitherto it has been assumed that the electron orbits are circular but, in general, periodic motion under the influence of a central force leads to elliptic orbits with the central body, e.g., the atomic nucleus, situated at a focus. The circle is, of course, a special case of an ellipse with major and minor axes having the same length. The momentum of a particle moving in an ellipse may be resolved into two components, one along the radius vector and the other at right angles; the latter then represents the angular momentum. Following the treatment used by Bohr, it may be supposed that this is quantized and can only have one of a definite series of values, given by the

expression $kh/2\pi$, where k is an integer which has been called the **azimuthal quantum number**. The momentum in the direction of the radius vector is also quantized, the value being determined in an analogous manner by means of a **radial quantum number**, represented by the symbol n_r . The combination of the two momenta leads to the equation

$$E = - \frac{2\pi^2 z^2 e^4 m}{h^2 (n_r + k)^2} \quad (78)$$

for the energy of the electron in a hydrogen-like atom, excluding the correction for nuclear motion; e , m , z and h have the same significance as before. Since n_r and k are whole numbers, it follows that their sum, which may be put equal to n , is also integral and (78) may be written

$$E_n = - \frac{2\pi^2 z^2 e^4 m}{h^2 n^2}, \quad (79)$$

which is identical with (63) deduced for circular orbits. The number n is now called the **principal quantum number**, and there are several possible values of k for each value of n , as may be readily shown. By the methods of classical dynamics it can be proved that n and k are in the same ratio as the axes of an elliptic path in which the electron is supposed to revolve; thus,

$$\frac{n}{k} = \frac{\text{length of major axis}}{\text{length of minor axis}}.$$

When n is equal to k the orbit must be circular, and a series of orbits of different eccentricities result for values of k equal to $n - 1$, $n - 2$,

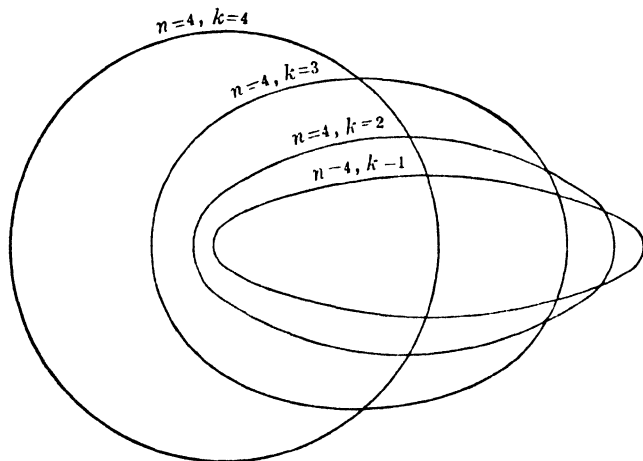


FIG. 13. Electron orbits

$n - 3$, ..., 1. According to the interpretation of the two quantum numbers just given, k cannot be zero since this would mean that the minor

axis of the ellipse is zero; in other words, the elliptical orbit of the electron would have degenerated into a straight line passing through the nucleus of the atom. If the electron is regarded as a particle moving in a definite path, this type of motion does not appear possible. It may be concluded, therefore, that there are n possible values of the azimuthal quantum number k , from n to 1, inclusive; thus if n is 4, the values of k can be 4, 3, 2 and 1, as seen in Fig. 13, which shows the four orbits with the ratio of axes 4:4, 4:3, 4:2 and 4:1, respectively.

It will be seen from (78) that the energies are apparently the same for all orbits having the same principal quantum number n , irrespective of the value of k , but Sommerfeld pointed out the necessity for a correction term. As a consequence of the movement of an electron in an elliptical orbit its actual velocity is not uniform. When the electron is closest to the nucleus—which, as indicated above, is situated at the focus of the ellipse—that is, in the “perihelion” position, its velocity is greater than when furthest from the nucleus, in the “aphelion” position. According to the theory of relativity, therefore, the effective mass of the electron will be different at different parts of its orbit (cf. p. 7), it being greatest at perihelion and least at aphelion. The curvature of the elliptical path is, therefore, not exactly the same in the two positions and the ellipse is thus not a complete one; the path does not return on itself, but rotates continuously in a circle, as shown in Fig. 14. The ellipse is said to have a “precessional” movement.* The energy of the electron in a hydrogenic atom in a particular state designated by the quantum numbers n and k is then found to be

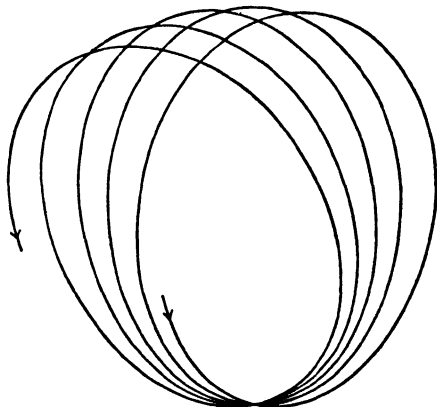


FIG. 14. Precessional movement of electron orbit

$$E = - \frac{2\pi^2 z^2 e^4 m}{n^2 h^2} \left[1 + \frac{\alpha^2 z^2}{n} \left(\frac{1}{k} - \frac{3}{4n} \right) \right], \quad (80)$$

excluding the correction for nuclear movement, where α , known as the **fine-structure constant**, is given by $2\pi e^2/hc$; this is equal to $1/137.0$ within the limits of accuracy of the accepted values of e and h . For a given principal quantum number n , therefore, the energy in any orbit

* The effect postulated by Sommerfeld is analogous to that observed in the orbit of Mercury, the planet nearest the sun, and first explained by Einstein's general theory of relativity.

depends to some extent on the azimuthal quantum number k , so that the frequency of the radiation emitted in the transition of an electron from a level designated by n_2 to the level n_1 will be slightly different according to the various possible k values. The small variations in energy will result in small energy, and hence frequency, differences, so that in the hydrogenic spectra a group of closely associated lines should appear, instead of a single line. It would seem, at first sight, that when n is large, and many k values are possible, the number of fine-structure lines should be very great. In practice this is not so, since k always changes by one unit only, that is, Δk is restricted to ± 1 . This **selection principle**, of which wave mechanics has given a theoretical interpretation, means that the actual number of lines is limited.

The frequency separations of a few fine-structure lines have been measured, and found to be in good agreement with the requirements of (80), in the Balmer series of hydrogen, and in the spectrum of ionized helium. The relativistic correction term in (80) involves z^4 , since z^2 appears both outside and inside the square brackets; consequently the frequency separation should depend on the fourth power of the effective charge for hydrogen-like atoms. For singly ionized helium, therefore, the frequency separation should be sixteen times as great as for hydrogen; the experimental observations are in agreement with this deduction.¹⁷

Weakness of the Bohr-Sommerfeld Theory.—Before giving further consideration to the applications of the Bohr theory, it is important to point out a weakness of the treatment already described. Although the Sommerfeld method of elliptical orbits gives the correct *total* (n) of possible azimuthal quantum numbers, the *actual values* are incorrect. Both experimental observation (see p. 63) and theoretical treatment based on wave mechanics show that the azimuthal quantum number can be zero, so that the values may be 0, 1, 2, \dots , $n - 1$, making a total of n possibilities. In order to avoid the risk of confusion the new azimuthal number is given the symbol l , where l is equal to $k - 1$, and an electron level, or orbit, is defined by the symbol n_l . Thus, if the principal quantum number n is 4, the possible sublevels are represented by 4_3 , 4_2 , 4_1 and 4_0 , for the azimuthal quantum number can be 3, 2, 1 and 0. On the basis of the pictorial model of Bohr and Sommerfeld these would represent increasing eccentricity of elliptical paths. In spite of its weakness this model may still be used as a basis for calculation; the azimuthal quantum number l will, however, be adopted in the subsequent discussion.

Penetrating Orbits.—In view of the ease with which the alkali metals become singly-charged positive ions, it is reasonable to assume that all the electrons except one are in closed, or completed, shells. The system is, therefore, hydrogen-like and all the inner electrons and the nucleus may be regarded as a single unit, the net charge on which is unity, i.e., $z = 1$, and around which the additional, i.e., optical, electron rotates. As a first

approximation it follows, therefore, that (67) should be applicable to the spectra of the alkali metals, and since $M/(m + M)$ is almost equal to unity, the term values should be given by R/n^2 , where R is $109,737 \text{ cm}^{-1}$. This relationship may be applied to calculate the *apparent* principal quantum numbers of the levels giving rise to the S , P , D and F terms of the alkali metals, to which reference has been made previously (p. 41). Some of the results obtained from the first four experimentally determined spectroscopic terms of the alkali metals are quoted as " n apparent" in Table 5. It is clear that " n apparent" is frequently not an integer, and consequently the terms would be more correctly represented as $R/(n - \delta)^2$, as suggested previously (p. 40). The true values of the principal quantum number n , which have been derived from various theoretical and experimental considerations, are given in the fourth column of Table 5; it

TABLE 5. CALCULATED APPARENT PRINCIPAL QUANTUM NUMBERS FOR THE ALKALI METALS

Atom	Term Series	n apparent	n	l
Lithium	S	1.59	2	0
	P	1.96	2	1
	D	3.00	3	2
	F	4.00	4	3
Sodium	S	1.63	3	0
	P	2.12	3	1
	D	2.99	3	2
	F	4.00	4	3
Potassium	S	1.77	4	0
	P	2.23	4	1
	D	2.85	3	2
	F	3.99	4	3
Rubidium	S	1.80	5	0
	P	2.28	5	1
	D	2.77	3	2
	F	3.99	4	3

will be seen that the discrepancies between true and apparent quantum numbers decrease in the order S , P , D , F of the term series, and in fact in the F series the difference is negligible. The discrepancies, sometimes called quantum defects, have been explained by the suggestion that the optical electron does not always remain completely outside the inner electron layers, so that it is not strictly correct to put z equal to unity, as postulated above. The greater the eccentricity of an electron orbit the more closely will the electron approach the nucleus in the course of its rotation (see Fig. 13), and so the more likely is it to penetrate the inner layers, and the greater will be the divergence from hydrogen-like behavior. It may be concluded from an examination of Table 5, therefore, that the penetration effect decreases in the order S , P , D , F , and this also

gives the order of decreasing eccentricity of orbit. In terms of the Bohr theory, the azimuthal quantum number k , that is $l + 1$, gives a measure of the orbital eccentricity, and hence it is possible to assign the values of l to the various terms, or energy levels, as shown in the last column of Table 5. The 2_1 , 3_2 and 4_3 orbits are circular, and hence the corresponding apparent and true values of n should be almost identical. It follows, therefore, that the S , P , D and F terms or states correspond in every case to l values of 0, 1, 2 and 3, respectively; this important generalization is now widely accepted.

Singly ionized atoms of the second group of the periodic classification, doubly ionized atoms of the third group, and so on, each have a single optical electron and so behave spectroscopically like alkali metal atoms. In these cases, however, the effective charge of the nucleus and of the inner electron shells is not unity, but it is in excess of this by the number of electrons lost when the ion is formed (cf. p. 39); thus for Mg^+ , Ca^+ , etc., the effective charge is 2, for Al^{++} it is 3, and for Si^{+++} it is 4, etc. The apparent quantum number is calculated from the fact that the term value should be Rz^2/n^2 , where z is the effective charge, and the results obtained in this way are quite analogous to those described above for the normal atoms of the alkali metals. Again the eccentricity of the orbits decreases in the order S , P , D , F , and the corresponding values of the azimuthal quantum number l are 0, 1, 2, 3.

If reference is made to p. 41, where the general expressions are given for the four main series of spectra, it will be seen that the following transitions occur in emission:

Series:	Sharp	Principal	Diffuse	Fundamental
Terms:	$S \rightarrow P$	$P \rightarrow S$	$D \rightarrow P$	$F \rightarrow D$
l values:	$0 \rightarrow 1$	$1 \rightarrow 0$	$2 \rightarrow 1$	$3 \rightarrow 2$

The corresponding values of l are given below the letter representing each term, or level, involved in the transition, and it is evident that the changes in l are always ± 1 , in agreement with the selection principle applicable to the atomic spectrum of hydrogen (p. 58).

A simplified energy level diagram for sodium is depicted in Fig. 15; those for the other alkali metals are similar, except for the difference in the principal quantum numbers of the optical electron. The integers indicate the various values of the principal quantum number n in each spectral state, represented by S , P , D or F . It may be noted that the P , D and F terms are actually doublets (p. 42), but this is not shown in the figure.

It is appropriate to inquire here why the hydrogen atom and singly ionized helium do not show S , P , D and F series in their spectra. The explanation is that these have only one electron and hence the question of the penetration of inner layers cannot arise. Apart from the relativity correction, therefore, the electronic energy is the same for all values of l , provided the principal quantum number is constant. When penetration

is possible, however, as for elements of higher atomic number, this is far from being the case.

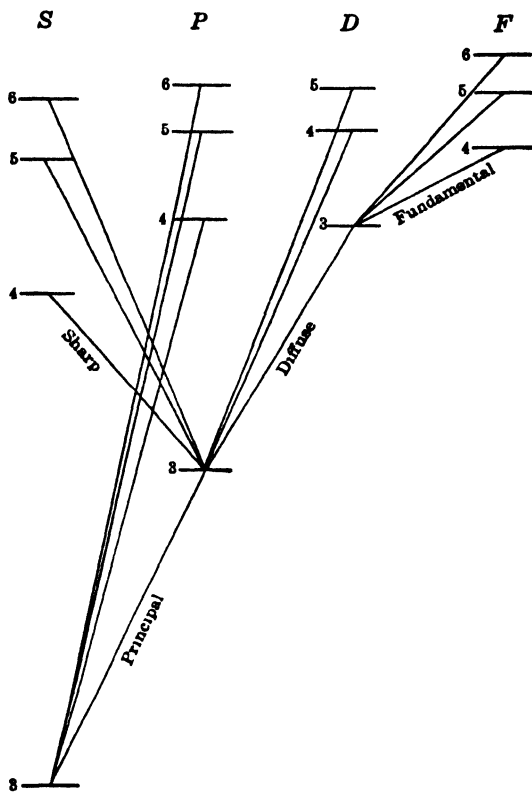


FIG.15. Formation of atomic spectra of sodium

Electron Spin.—When describing the nature of spectra it was pointed out that many lines are actually multiplets consisting of groups of two, three or more lines close together. The interpretation of this fact proved a matter of difficulty until G. E. Uhlenbeck and S. Goudsmit (1925) put forward the suggestion that an electron rotates about its own axis, that is to say, it spins, and consequently contributes to the total angular momentum. This contribution is supposed to be quantized, the **spin quantum number**, represented by s , being capable of only two values, viz., $+\frac{1}{2}$ or $-\frac{1}{2}$, in terms of the usual $h/2\pi$ units of angular momentum. The alternative positive and negative signs may be assumed to represent the electron spinning in one direction or the other. The resultant of the spin and azimuthal contributions to the angular momentum is obtained in terms of $h/2\pi$ units by adding l and s , so that

$$j = l + s, \quad (81)$$

where j is sometimes called the **inner quantum number**, although this is an unsatisfactory name; the resultant angular momentum would then be $j\hbar/2\pi$.* Since s can be $+\frac{1}{2}$ or $-\frac{1}{2}$, it follows that for every value of l there are two values of j , viz.,

$$j_1 = l + \frac{1}{2} \quad \text{and} \quad j_2 = l - \frac{1}{2},$$

except when l is 0, for then the two j values are identical, viz., $+\frac{1}{2}$ and $-\frac{1}{2}$, for it is the *numerical* value only of j which determines the momentum. Every l level, except that for which l is 0, is consequently split into two levels with different energies. The possibility of fine structure is evident, although the number of lines is limited by the selection principle which may be written in the form $\Delta j = 0$ or ± 1 .

For an electron in an S level the value of l is 0; hence it is a singlet level † with $j = \frac{1}{2}$, and all S terms are, therefore, singlets. In the P levels, l is 1 and so the corresponding j values are $\frac{3}{2}$ and $\frac{1}{2}$; in the D levels, where l is 2, they are $\frac{5}{2}$ and $\frac{3}{2}$; and finally, in the F levels, l is 3 and so the j values are $\frac{7}{2}$ and $\frac{5}{2}$. The doublets of the Sharp series of spectra (cf. p. 42) may then be represented by the expression

$$\bar{\nu}_1 = aP_{1/2} - nS \quad \text{and} \quad \bar{\nu}_2 = aP_{3/2} - nS,$$

the S term being a singlet and the P term a doublet; the j value in each case is indicated by a suffix to the term symbol. The frequency separation of the lines of all the Sharp doublets is $aP_{1/2} - aP_{3/2}$, which is clearly constant, as found experimentally. For the Principal series the doublets are

$$\bar{\nu}_1 = aS - nP_{1/2} \quad \text{and} \quad \bar{\nu}_2 = aS - nP_{3/2},$$

so that the frequency separation $nP_{3/2} - nP_{1/2}$ is not constant, but increases as the value of the principal quantum number n increases. The lines of the Diffuse and Fundamental series consist of triplets, and not quadruplets as might be expected, because of the selection principle, $\Delta j = 0$ or ± 1 ; thus,

Diffuse series:

$$\bar{\nu}_1 = aP_{1/2} - nD_{3/2}, \quad \bar{\nu}_2 = aP_{3/2} - nD_{3/2}, \quad \text{and} \quad \bar{\nu}_3 = aP_{3/2} - nD_{5/2}.$$

Fundamental series:

$$\bar{\nu}_1 = aD_{3/2} - nF_{5/2}, \quad \bar{\nu}_2 = aD_{5/2} - nF_{5/2}, \quad \text{and} \quad \bar{\nu}_3 = aD_{5/2} - nF_{7/2}.$$

Actually the triplets are not always observed because some of the frequencies are so close together that the lines are virtually inseparable, and so they often appear broad and diffuse, as in the Diffuse series.

One important fact may be mentioned in conclusion: the observations on spectral multiplets clearly require that the S levels should be singlets, and so

* It is important to note that according to modern theory, based on quantum mechanics, the orbital angular momentum $l\hbar/2\pi$ is replaced by $\sqrt{l(l+1)}\hbar/2\pi$; similarly the spin momentum is $\sqrt{s(s+1)}\hbar/2\pi$, instead of $s\hbar/2\pi$. The resultant angular momentum is then $\sqrt{j(j+1)}\hbar/2\pi$, and not as given above.

† The adjectives singlet, doublet, triplet, etc., and multiplet are used for energy levels or spectroscopic terms as well as for the actual lines in the spectrum.

the azimuthal quantum number must be zero. This is the first argument encountered against the use of the original k values of Sommerfeld, since these did not permit of a zero azimuthal quantum number (cf. p. 57).

Magnetic Quantum Numbers.—In a strong magnetic field the lines of the spectrum are split up; this is known as the Zeeman effect (P. Zeeman, 1896). If the spectrum is examined in a direction at right angles to the magnetic field, each line is seen to be split into three, the middle one being polarized in a direction parallel to the field, and the outer lines at right angles to it; when the lines are observed in the direction parallel to the field each is found to be split into two components, which are circularly polarized (see p. 597) in opposite directions. To explain these phenomena by the Bohr theory it was necessary to postulate an additional **magnetic quantum number**, m , where $mh/2\pi$ represents the component of the total angular momentum of the electron in the direction of the field. If this component is quantized, as suggested, it means that the total angular momentum can only have a finite number of possible inclinations with respect to the direction of the magnetic field; this may be represented pictorially in Fig. 16, for the case in which the azimuthal quantum number l is 3. The possible inclinations, seven in all, of the total angular momentum are shown by arrows, the corresponding m values being $+3, +2, +1, 0, -1, -2$ and -3 ; in general, therefore, there should be $2l + 1$ possible values of m for every value of l . The introduction of the spin quantum number, which was a later development, necessitated a slight change, the possible m values being now $2j + 1$, thus: $0, \pm 1, \pm 2, \dots, \pm j$ for each j value. The precessional motion of the electron orbit in a magnetic field will contribute energy to the electron, so that in a strong field each electron level is split into $2j + 1$ sublevels, there being $2j + 1$ possible settings relative to the direction of the magnetic field. This evidently means that the spectral lines will also split in a magnetic field, but the number of components is limited by the selection rule $\Delta m = 0$ or ± 1 perpendicular to the field, and $\Delta m = \pm 1$ parallel to the field. The possible transitions are thus:

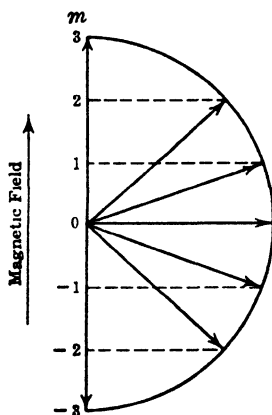


FIG. 16. Magnetic quantum numbers

Perpendicular to the magnetic field	$m \rightarrow m + 1$	$m \rightarrow m$	$m \rightarrow m - 1$
Parallel to the magnetic field	$m \rightarrow m + 1$		$m \rightarrow m - 1$

If the energy changes in these transitions are independent of the actual value of m , each line will be split into three perpendicular and two parallel to the direction of the magnetic field. It may be mentioned that the situation is not quite so simple as that described here. In magnetic

fields of fairly low strength an anomalous Zeeman effect is observed, in which more lines are found than are required by the normal effect, and at intermediate field strengths another type of splitting, known as the Paschen-Back effect (F. Paschen and E. Back, 1921) can be detected. These phenomena can be explained by means of wave mechanics and the concept of electron spin, but they need not be considered further in this book.

Quantization of Direction.—The existence of $2j + 1$ magnetic quantum numbers for each value of j implies that there is a possibility of $2j + 1$ different orientations in a magnetic field, as indicated by the arrows in Fig. 16; in other words, there should be a quantization of direction. Direct experimental proof of this expectation was provided by the important work of O. Stern and W. Gerlach (1921). A stream of atoms, produced from a heated wire or other source and narrowed down into a thin beam by passage through slits, was passed through an inhomogeneous magnetic field obtained by means of a powerful magnet MM with pole pieces of different shapes, as shown in Fig. 17, and then allowed to fall on a plate P . Every electron in an atom

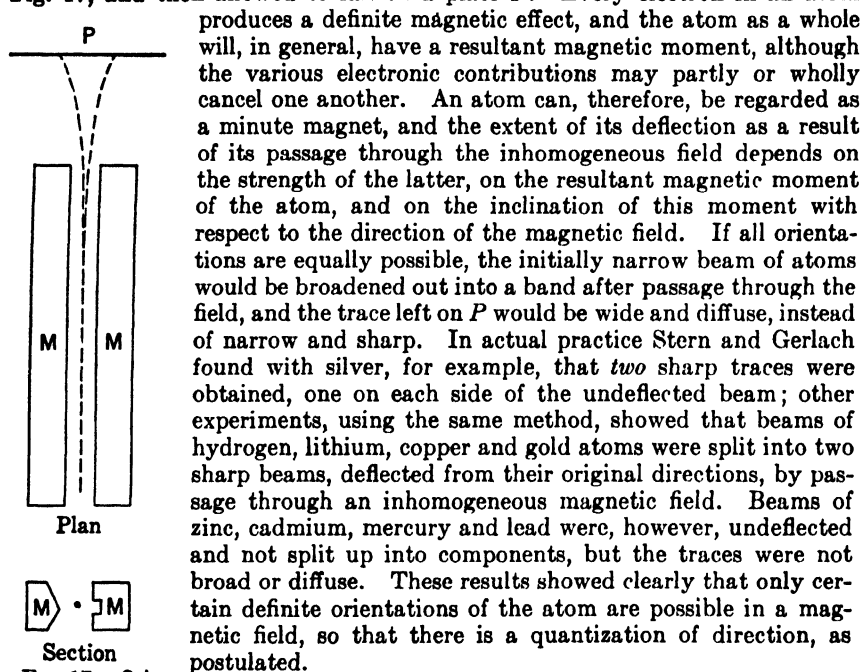


FIG. 17. Orientation of atoms in a magnetic field (Stern and Gerlach)

produces a definite magnetic effect, and the atom as a whole will, in general, have a resultant magnetic moment, although the various electronic contributions may partly or wholly cancel one another. An atom can, therefore, be regarded as a minute magnet, and the extent of its deflection as a result of its passage through the inhomogeneous field depends on the strength of the latter, on the resultant magnetic moment of the atom, and on the inclination of this moment with respect to the direction of the magnetic field. If all orientations are equally possible, the initially narrow beam of atoms would be broadened out into a band after passage through the field, and the trace left on P would be wide and diffuse, instead of narrow and sharp. In actual practice Stern and Gerlach found with silver, for example, that *two* sharp traces were obtained, one on each side of the undeflected beam; other experiments, using the same method, showed that beams of hydrogen, lithium, copper and gold atoms were split into two sharp beams, deflected from their original directions, by passage through an inhomogeneous magnetic field. Beams of zinc, cadmium, mercury and lead were, however, undeflected and not split up into components, but the traces were not broad or diffuse. These results showed clearly that only certain definite orientations of the atom are possible in a magnetic field, so that there is a quantization of direction, as postulated.

A beam of hydrogen atoms is split into two in a magnetic field, so that for atomic hydrogen $2j + 1$ is 2, and hence j is $\frac{1}{2}$; this is only possible if l is 0 for the one electron of hydrogen. There is thus a further argument against the classical view that the azimuthal quantum number cannot be zero. For the other atoms mentioned which also have two possible orientations, the resultant $2j + 1$ value for all the electrons must be 2; when there is no splitting, the

resultant is 1. The relation between the resultant and individual quantum numbers of the electrons will be considered shortly.¹⁹

The Magnetron.—It can be shown by the methods of classical electrodynamics that the magnetic moment m_μ due to an electron rotating in an orbit is given by

$$m_\mu = \frac{ep}{2mc}, \quad (82)$$

c being the velocity of light, e the electronic charge, p the total mechanical angular momentum of the electron, and m its mass. The momentum will be quantized in $h/2\pi$ units, as postulated by Bohr, so that the unit of magnetic moment (m_{μ_0}) may be obtained by assuming the quantum number to be unity, and setting p equal to $h/2\pi$, thus

$$m_{\mu_0} = \frac{eh}{4\pi mc}. \quad (83)$$

The quantity m_{μ_0} is called the **Bohr magneton**; it is equal to 9.273×10^{-21} erg gauss⁻¹, and represents the fundamental unit of magnetic moment per atom. To obtain the moment per g. atom (M_{μ_0}) it is necessary to multiply by the Avogadro number, 6.023×10^{23} , giving 5585 erg gauss⁻¹. By the simple quantum theory the magnetic moments of atoms should be integral multiples of m_{μ_0} , but the methods of wave mechanics have shown that the total angular momentum of an electron is $\sqrt{j(j+1)}h/2\pi$ (see footnote, p. 62), so that actual moments should be $\sqrt{j(j+1)}$ Bohr magnetons. In addition it is found necessary to introduce a "splitting factor" g (A. Landé, 1923), so that the moment m_μ of an atom with *one optical electron* is determined by

$$m_\mu = \sqrt{j(j+1)}gm_{\mu_0}, \quad (84)$$

where the splitting factor is defined as

$$g = 1 + \frac{j(j+1) + s(s+1) - l(l+1)}{2j(j+1)}, \quad (85)$$

the values of the quantum numbers j , l and s being for the atom in its normal state of lowest energy, that is, in the ground state. Equation (84) applies to hydrogen-like and alkali metal-like atoms, but when more than one orbital electron is involved the resultant quantum numbers J , L and S , defined later (p. 67), replace j , l and s , respectively.

It will be shown in Chapter VIII (p. 616) that the atomic paramagnetic susceptibility χ_μ , which is the experimentally determinable quantity, is related to the magnetic momentum m_μ by the expression

$$\chi_\mu = \frac{N}{3kT} m_\mu^2, \quad (86)$$

where N is the Avogadro number and k is the Boltzmann constant, i.e., the gas constant per single molecule (p. 264); combination of (84) and (86), therefore, gives

$$\chi_\mu = \frac{Nj(j+1)g^2}{3kT} m_{\mu_0}^2. \quad (87)$$

A more exact treatment, based on the quantum theory of paramagnetism (J. H.

Van Vleck and Miss A. Frank, 1929), indicates the necessity of adding another term, $N\alpha$, to the right-hand side of (87), where α depends on the quantum numbers of the atom; for most elements the correction term is negligible, although it is of considerable importance for certain rare-earth ions (p. 91).

Two or More Optical Electrons.—The discussion so far has been restricted to spectra of atoms or ions possessing only one optical electron. When there are two or more such electrons the spectra are more complicated, although the Ritz combination principle is still applicable. The postulation of four quantum numbers for an electron permits a relatively simple interpretation of the spectra of the alkali metals and the influence of a magnetic field, and the same concepts can be applied to the atomic spectra of other elements. The metals of the alkaline-earth group have two optical electrons, with the result that the spectra are not so simple as those of the first group. Nevertheless, it has been possible by analysis to show that the atomic spectra of the former, as well as of zinc, cadmium and mercury, are characterized by two almost independent term systems, one being a singlet system and the other a triplet system. In each of the systems it is generally feasible to identify the four types of spectra, viz., Sharp, Principal, Diffuse and Fundamental (Bergmann), corresponding to those observed for elements having only one optical electron, but in addition a number of lines representing transitions from one system to another are observed. The arc spectra of the elements of the third group of the periodic table, e.g., aluminum, scandium, gallium and thallium, require that the levels should consist of doublets and quartets, forming two virtually independent systems, whereas in the fourth group singlet and triplet systems have been detected for carbon and silicon, and in addition there are quintet systems in the spectra of titanium and zirconium. The spectra of elements in a number of other groups of the periodic system have been analyzed and important regularities observed. Some of these will be evident from Table 6, which gives the multiplicities of the levels represented in the arc spectra of the elements from potassium to zinc.

TABLE 6. OBSERVED TERM MULTIPLICITIES IN ATOMIC (ARC) SPECTRA

K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn
2	1	2	1	2	(1)	(2)	(1)	2	1	2	1
	3	4	3	4	3	4	3	4	3	4	3
			5	6	5	6	5	6	5		
					7	8	7				

It is evident that there is a definite alternation in term type as the atomic number increases; the results may be expressed by the **alternation law** in the following form. *The spectra of the successive groups of the periodic table of elements are characterized by an alternation of odd and even term multiplicities; the elements of odd valence have even multiplets, and vice versa.* It should be noted that in Table 6 some of the multiplicity values are in heavier type; these are for the terms giving rise to the most

intense spectra, and hence are regarded as being associated with the normal (ground) state of the atom when it is in its lowest energy level. According to the "maximum multiplicity rule" of F. Hund (1925), the ground term is associated with the largest multiplicity possible for the available electrons. This is important in connection with the elucidation of the electronic configurations of atoms.

Resultant Quantum Numbers.—The interpretation of the multiplet terms in the spectra of elements with more than one optical electron, by the aid of the quantum numbers of these electrons, depends on a knowledge of how the two forms of angular momentum, viz., orbital and spin, represented by the quantum numbers l and s , respectively, couple, or combine, with one another. The most frequent type of coupling is that described by H. N. Russell and F. A. Saunders (1925), in which it is assumed that all the orbital angular momenta of the different electrons couple together to give a total determined by a resultant quantum number L ; this must be integral, and equal to the vectorial sum of the l values for all the electrons. The summation is simplified by the fact that electrons in closed shells (p. 81) do not contribute to L , so that only the relatively few optical electrons need to be considered even for elements of high atomic weight. For two such electrons, for example, having azimuthal quantum numbers l_1 and l_2 , the possible values of L , representing different ways of combining l_1 and l_2 vectorially, are:

$$L = l_1 + l_2, \quad l_1 + l_2 - 1, \quad l_1 + l_2 - 2, \quad \dots, \quad l_1 - l_2.$$

Similarly, angular momenta due to spin couple together, so that the resultant is determined by the quantum number S ,* which is equal to the algebraic sum of the s values for the separate electrons. Further, just as l and s values combine to give j for a single electron, so the possible resultant angular momenta are determined by a series of J values obtained by different vectorial combinations of L and S . Two simple examples, with $L = 1$ and $S = 1$, and with $L = 2$ and $S = 1$, are shown in Fig. 18, when it is seen that three values of J are possible in each case. It can be shown that, in general, the possible values of J are:

$$J = L + S, \quad L + S - 1, \quad L + S - 2, \quad \dots, \quad |L - S|, \dagger$$

so that there will be a total of $2S + 1$ values of J , for each L value as given above. The resultant angular momentum is then $\sqrt{J(J + 1)}h/2\pi$, according to the particular value of J in any state of the atom.

Explanation of Term Multiplicities.—It is now possible to see how the assumption of Russell-Saunders coupling accounts for the observed term multiplicities. For example, if there are two optical electrons, then the spins may be

* Care should be taken to avoid confusion between S used here for the resultant spin quantum number and as already employed for a term symbol.

† The symbol $|L - S|$ implies that only positive (or zero) values of J are significant when $S > L$.

in opposite directions, that is antiparallel, when $S = 0$, or they may be in the same sense, that is parallel, when S is 1, since the individual s values for each electron can only be $+\frac{1}{2}$ or $-\frac{1}{2}$. If S is 0, then there will be only one J value, equal to L ; but if S is 1, three values are possible, namely $L + 1$, L and $L - 1$ (cf. Fig. 18). Just as the j values determine the multiplicity of each l level when there is only one optical electron, so the J values indicate the number of multiplets to be expected for each term in the spectrum of an atom with several

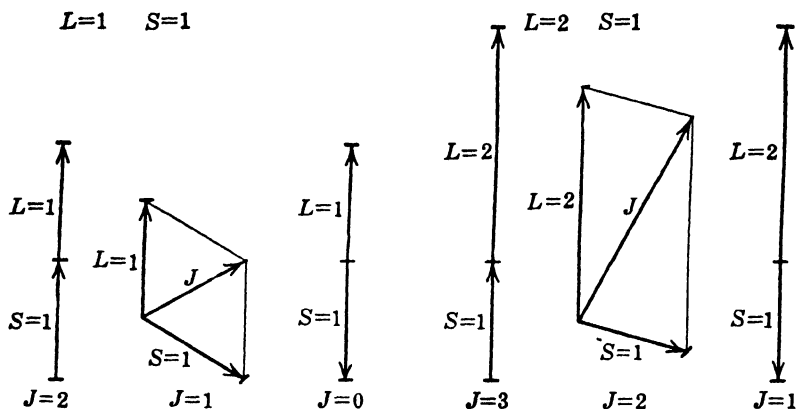


FIG. 18. Russell-Saunders coupling

optical electrons. For the elements of the second group of the periodic system, therefore, there should be, as actually found, two term systems consisting of singlets and triplets, corresponding to the one and three possible values of J , respectively, for each L value.

When there are three optical electrons, as in the arc spectra of elements of the third group, the resultant spin vector can be $\frac{1}{2}$ or $\frac{3}{2}$, made up as follows: $\pm \frac{1}{2} \mp \frac{1}{2} \pm \frac{1}{2}$ and $\pm \frac{1}{2} \pm \frac{1}{2} \pm \frac{1}{2}$, so that for each L value the possible J values are $L + \frac{1}{2}$ and $L - \frac{1}{2}$, and $L + \frac{3}{2}$, $L + \frac{1}{2}$, $L - \frac{1}{2}$ and $L - \frac{3}{2}$, respectively. There will be, consequently, two sets of terms, one consisting of doublets and the other of quadruplets. In the fourth group of the periodic table, S can be 0, 1 or 2, so that there should be three sets of terms; namely, singlets, triplets and quintuplets, for, as already mentioned, there are $2S + 1$ values of J for every value of L . This conclusion, like those reached for the earlier groups, is in agreement with the results obtained from a study of the arc spectra of the various elements. The observed multiplicities of the levels in the subsequent groups can be accounted for in an analogous manner, and so the results recorded in Table 6, and the alternation law, can be derived theoretically. It will be noted that Hund's maximum multiplicity rule implies that in the ground state the maximum possible number of electrons have unpaired spins.

***jj*-Coupling.**—Although the *LS* (Russell-Saunders) type of coupling, in which the interaction of the individual l vectors and of the s vectors is strong, is applicable to the majority of elements, in some cases the s and l , i.e., "spin-orbit," interaction for each electron is stronger. This leads to what is called *jj*-coupling. The l and s values of each electron are combined to give a j value, and the coupling of these for all the electrons in the atom determines J . In this type of coupling the J value is consequently definite, but L and S are

not. The number of terms remains the same as for LS coupling, but the selection rules are different. Since the l and s interactions are variable, from atom to atom, it is to be expected that LS and jj couplings are extremes, with the possibility of intermediate cases. Actually, such a gradual transition has been observed with increasing atomic weight.

Symbols for Atomic States.—In order to represent more completely the electronic state of an atom, a scheme based on the use of spectral term symbols, which is applicable to all atoms, was introduced by H. N. Russell and F. A. Saunders (1925). The term letter S, P, D, F, G, H , etc., is used according as the value of L in the particular atomic state is 0, 1, 2, 3, 4, 5, etc.; it is preceded by a superscript representing the multiplicity of the term, i.e., $2S + 1$, and followed by a subscript giving the corresponding J value. For example, with an electronic coupling giving $L = 2$ and $S = 1$, the multiplicity of each term is 3, and the term symbols would be ${}^3D_3, {}^3D_2$ and 3D_1 , according as J is 3, 2 and 1, respectively, each symbol representing one component of the triplet term or energy level. In order to identify a particular electronic level it is necessary to give also the principal quantum number, which precedes the symbol, e.g., 5^3D_3 , where n is 5; this is read as "five-triplet- D -three," and represents an electronic state of the atom wherein the optical electrons have $n = 5, L = 2, S = 1$ and $J = 3$.

When there is only one optical electron, L is equal to l , and J to j , so that the Russell-Saunders representation is the same as that previously employed, with the addition of a superscript indicating the multiplicity. The P, D and F levels of the alkali metals are doublets (p. 42) so that the appropriate term symbols are $n^2P_{1/2}$ and $n^2P_{3/2}$, etc. Theoretically, the S terms are also doublets, since the multiplicity is $2s + 1$, and s is $\frac{1}{2}$. As already seen, however, the two levels are identical because the j values are $+\frac{1}{2}$ and $-\frac{1}{2}$, which are indistinguishable as far as the total angular momentum is concerned. It is, nevertheless, the practice to employ the doublet superscript, viz., $n^2S_{1/2}$.

For an element with two valence electrons there are two series of terms, singlets and triplets, as shown above, with J values equal to L , and to $L + 1, L$ and $L - 1$, respectively. The various possible terms are then ${}^1S_0, {}^1P_1, {}^1D_2$ and 1F_3 for the singlets, and ${}^3S_1, {}^3P_2, {}^3P_1, {}^3P_0, {}^3D_3, {}^3D_2, {}^3D_1, {}^3F_4, {}^3F_3$ and 3F_2 for the triplets. The 3S_1 term is actually a singlet, since the three J values are identical, but the triplet superscript is retained.

It may be mentioned that for certain atoms not all the terms theoretically possible are actually observed. Owing to the operation of an important principle to be considered shortly (p. 80) there are certain limitations to the distribution of quantum numbers, and hence some of the levels are said to be "forbidden."²⁰

X-Ray Spectra.—Reference has been made earlier (p. 25) to the emission by each element of characteristic X-rays consisting of two or more groups of rays of definite frequencies; these groups are designated the K -, L -, M -, N - and O -series, in order of increasing wave length and

of decreasing hardness or penetrating power. With the lighter elements the K - and L -series only can be excited, but the others are obtained with increasing atomic number, or atomic weight. The spectral lines of each series are not simple but, as with optical spectra, consist of closely spaced multiplets generally represented by the symbols K_α , K_β , L_α , L_β , L_γ , etc. On the whole, however, X-ray spectra are very much simpler than optical spectra appearing in the ultraviolet and visible regions, and there are other differences. For example, in optical spectra successive lines can be excited by gradually increasing the excitation energy employed (p. 36), but with X-rays no emission occurs until the excited electron acquires an energy corresponding to the series limit, when all the lines in the particular series appear simultaneously. A further difference is seen in the fact that optical spectra are similar in type for elements in any one group of the periodic table, but vary considerably from one group to the next; on the other hand, the X-ray spectra of all elements are analogous, the frequencies of corresponding lines changing regularly as the atomic number increases. It has been recorded (p. 25) that Moseley showed the square root of the frequencies of the K_α lines to be proportional to $Z - \sigma$, where Z is the atomic number, and σ is practically equal to unity, and it is possible to write, for this series,

$$\bar{\nu} = \frac{3}{4}R(Z - 1)^2, \quad (88)$$

where R is the same Rydberg constant as occurs in the equations for the frequencies of optical spectra. The equation (88) may be rewritten in the form

$$\bar{\nu} = R(Z - \sigma)^2 \left(\frac{1}{1^2} - \frac{1}{2^2} \right), \quad (88a)$$

where σ is 1; similarly, the frequencies of the lines of the L_α series for different elements may be represented by

$$\bar{\nu} = R(Z - \sigma)^2 \left(\frac{1}{2^2} - \frac{1}{3^2} \right), \quad (89)$$

where σ is approximately 7.4 for the whole series. The frequencies of the characteristic X-ray lines may thus be written as the difference of two terms, as with optical spectra, and a principle analogous to the Ritz combination principle is applicable.

Origin of X-Ray Spectra.—The interpretation of X-ray spectra in terms of the Bohr theory of electronic levels was first proposed by W. Kossel (1920). It is supposed that the electrons in an atom are accommodated, as has been frequently implied, in a series of layers or shells, the first being called the K -shell, the second the L -shell, and so on. These shells are now identified with groups of electrons possessing the same principal quantum number; thus, all electrons for which n is 1 are in the K -shell, those for which n is 2 are in the L -shell, etc. The electrons in the same shell have approximately the same energy and so are bound more or

less equally strongly to the nucleus; the strength of binding may be supposed to fall off with increasing distance of the shell, or layer, from the nucleus, so that it diminishes in the order K, L, M , etc. For a K -line to be excited, it is postulated that the atom absorbs sufficient energy to eject an electron from the K -shell. Since the L -, M -, N -, etc., shells may be fully occupied, the electron may have to be moved to a still higher level, and the requisite energy may be considerable. The vacancy in the K -level can now be filled by an electron moving in from the L -, M -, N -, etc. layers, the energy difference being emitted in the form of the K -series of X-rays (see Fig. 19). This accounts for the observations that all K -lines are excited simultaneously, and that it is not possible to produce any particular

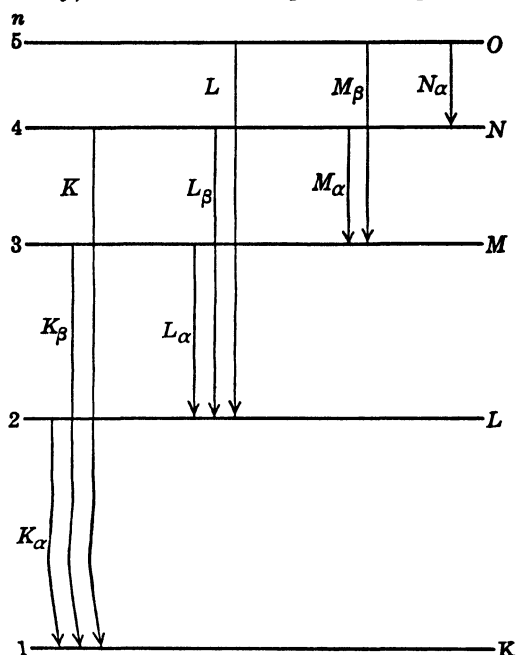


FIG. 19. Formation of characteristic X-rays

line by supplying only an equivalent amount of energy, as is possible with optical spectra. The K_α -lines represent transitions from a level of principal quantum number 2 to one in which it is 1, and so the nature of the terms in (88a) is explicable. In a similar manner, it can be seen that the ejection of an electron from the L level results in the excitation of the L -series of X-rays, when an electron from an M -, N -, etc. layer falls into the L -shell. The form of (89) suggests that the L_α -line results from transitions between M -, for which $n = 3$, and L -, wherein $n = 2$, levels.

Since the energy of binding of an electron falls off in the order K, L, M, N, O , this will also represent the order for the ease of excitation of the corresponding X-ray series. As a result of the increasing ease of absorp-

tion with increasing wave length, however, the higher series are difficult to detect and study. It will be observed, further, that for light atoms, which normally have electrons restricted to the K -, L - and M -shells, the K - and L -series of X-rays only can be excited, but with increasing atomic number the X-ray series of longer wave length can be obtained. It should be remembered that the small differences of energy of the electrons in each shell, with the same principal quantum number, will result in a fine structure of the X-ray lines, as is actually observed; this point, however, need not be discussed further.²¹

The Screening Constant.—In (88a) and (89), the effective charge is represented by $Z - \sigma$; the quantity σ , which is approximately constant throughout any given series, is called the **screening constant**. When lines of the K -series are produced, it may be supposed that any electrons outside the K -shell have little influence on the effective charge. It will be seen later (p. 83) that the complete K -shell probably contains two electrons, and when one has been removed in the process of excitation, one only is left to screen the nucleus. It is thus evident that for the K_α -series, at least, the screening constant σ should be approximately unity, and the effective charge is $Z - 1$, as found by experiment. The plot of $\sqrt{\nu}/R$ against Z is very close to a straight line for the K_α -lines of all elements from boron onward (cf. Fig. 8), so that the screening factor σ must be constant throughout. In the L -series, it is to be expected that the screening constant will be greater, for the nucleus is screened by the K -shell and also by the L -shell, which can contain a maximum of eight electrons. The value of σ is found to be approximately 7.4; it does not remain constant throughout the whole series, for the plot of $\sqrt{\nu}/R$ for the L -lines against the atomic number, starting with sodium, is only approximately a straight line, although there are no significant breaks. In the M - and higher series, abnormalities are observed in connection with the screening effect; for the groups of elements of atomic number 21 to 29, 39 to 47, and 57 to 71, inclusive, the slope of the plot of $\sqrt{\nu}/R$ against Z for a given series of X-rays is less than for the other elements. In some instances $\sqrt{\nu}/R$ does not change appreciably, showing that $Z - \sigma$ is remaining constant instead of increasing, in spite of the regular increase in Z , the atomic number. This result implies that for elements behaving in this manner the screening constant increases with the atomic number, so that one additional electron is apparently inserted in an inner shell for each increase of unity in the atomic number. The importance of this conclusion will be considered below (p. 87).

Objections to the Bohr Atom.—It is true that the concepts introduced by Bohr have achieved remarkable successes in many directions, but there are a number of fundamental weaknesses to which attention must be drawn. The methods of classical physics are used to determine the path of an electron, but these are immediately discarded when the existence of stationary levels and quantization of momentum are postu-

lated; this inconsistency was only justified by its giving correct results. There are, in addition, the points already mentioned: first, that the Bohr-Sommerfeld treatment leads to an incorrect assignment of azimuthal quantum numbers, and second, that the exact definition of position, or orbit, and momentum is contrary to the Heisenberg uncertainty principle, now regarded as one of the essential foundations of physics. Further, the use of planar orbits implies, at least for hydrogen, a flat atom, which appears to be contrary to the result of experiment. Apart from certain aspects, that need not be considered here, the Bohr theory gives a reasonably satisfactory quantitative interpretation of the spectra of atomic hydrogen and of ionized helium, but it fails completely for normal helium. The work of Bohr has brought out, however, the essential points of the quantization of momentum and the existence of definite energy levels, although his picture of the atom probably does not represent any physical reality.

WAVE MECHANICS

Wave Mechanics.—Two promising alternative procedures for investigating the mechanics of the atom have been developed in recent years: one of these is the matrix mechanics of W. Heisenberg (1925), which is purely mathematical and involves no atomic model, and the other is the wave mechanics of E. Schrödinger (1926), based on the suggestions (see p. 14) of L. de Broglie (1924), which has a physical aspect although, it must be admitted, a somewhat vague one. Since the ideas underlying the latter treatment have become important in physical chemistry, and are not difficult to grasp, they will be considered briefly. According to the views of de Broglie, every particle is associated with a wave (p. 15), and the wave length (λ) is related to the mechanical momentum (mv) by (20); thus,

$$\lambda = h/mv,$$

where h is the Planck constant. When an electron moves in a circular orbit round a nucleus, it may be supposed that the associated wave train extends round the circle, and if the wave is to remain continually in phase, the circumference of the circle must be an integral (n) number of wave lengths (Fig. 20, A): that is,

$$2\pi r = n\lambda = nh/mv; \quad (90)$$

$$\therefore mvr = \frac{nh}{2\pi}, \quad (91)$$

where the symbol m represents the electronic mass. Since mvr is the angular momentum, it follows immediately from the wave concept that only those circular orbits are permitted for which the angular momentum is an integral number of $h/2\pi$ units; this is exactly Bohr's postulate (p. 44), for which there was originally no theoretical basis. The above

argument is not rigid, for the electronic wave system has been assumed to be two-dimensional only, and in any event has no reality, but the result is of interest.

The Wave Equation.—The development of the Schrödinger wave equation may be considered by examining, first, the simplest type of wave motion, namely the vibrations of a stretched string. If w is the amplitude at any point whose coordinate is x , at time t , the appropriate equation for the wave motion is

$$\frac{\partial^2 w}{\partial x^2} = \frac{1}{u^2} \cdot \frac{\partial^2 w}{\partial t^2}, \quad (92)$$

where u is the velocity of propagation of the waves. This differential equation may be solved by separating the variables; thus, w may be expressed in the form

$$w = f(x)g(t), \quad (93)$$

where $f(x)$ is a function of the coordinate x only, while $g(t)$ is a function of

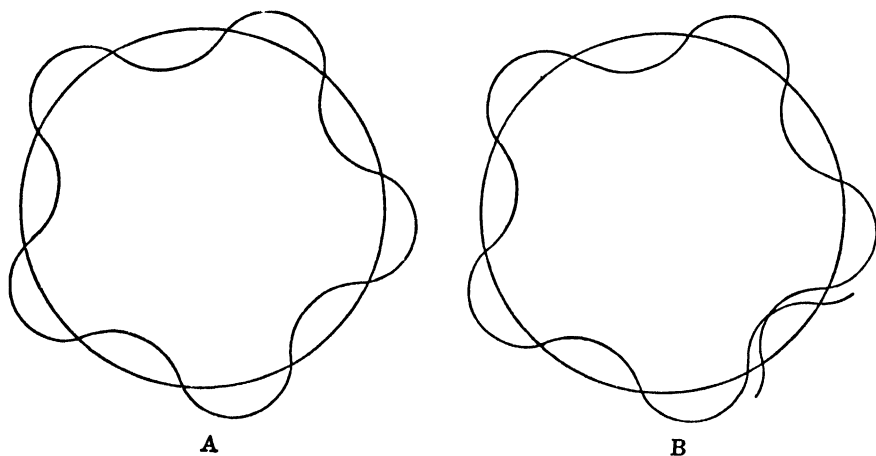


FIG. 20. Electron waves and orbits

the time only. For the motion of standing waves, such as occur in a stretched string, it is known that $g(t)$ may be represented by

$$g(t) = A \sin 2\pi\nu t,$$

where ν is the vibration frequency and A is a constant, the maximum amplitude. The equation for w may then be written as

$$w = f(x)A \sin 2\pi\nu t,$$

and hence

$$\frac{\partial^2 w}{\partial t^2} = -f(x)4\pi^2\nu^2(A \sin 2\pi\nu t) = -4\pi^2\nu^2 f(x)g(t). \quad (94)$$

From (93) it follows that

$$\frac{\partial^2 w}{\partial x^2} = \frac{\partial^2 f(x)}{\partial x^2} g(t),$$

and combination with (92) and (94) gives

$$\frac{\partial^2 f(x)}{\partial x^2} = -\frac{4\pi^2 \nu^2}{u^2} f(x). \quad (95)$$

The frequency of the vibrations ν and the velocity u are related by the expression $u = \lambda \nu$, where λ is the corresponding wave length; hence, from (95),

$$\frac{\partial^2 f(x)}{\partial x^2} = -\frac{4\pi^2}{\lambda^2} f(x). \quad (96)$$

This equation for wave motion in one direction may be extended to movement in three directions, represented by the coordinates x , y and z . If $f(x)$ for one coordinate is replaced by $\psi(x, y, z)$, which is the amplitude function for three coordinates, (96) takes the form

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} = -\frac{4\pi^2}{\lambda^2} \psi, \quad (97)$$

where ψ has been written, for the sake of brevity, in place of $\psi(x, y, z)$. Utilizing the symbol ∇^2 for the differential (Laplacian) operator, i.e.,

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2},$$

(97) can be written more concisely as

$$\nabla^2 \psi = -\frac{4\pi^2}{\lambda^2} \psi. \quad (98)$$

The fundamental postulate of the Schrödinger treatment is that an equation of the type of (98) may be applied to all particles, including electrons, atoms and photons. Upon introducing, as a further postulate, the de Broglie relationship $\lambda = h/mv$ into (98) this becomes

$$\nabla^2 \psi = -\frac{4\pi^2 m^2 v^2}{h^2} \psi, \quad (99)$$

where m is the mass and v the velocity of the particle.* The kinetic energy of the latter, which is the difference between the total energy E and the potential energy U , is equal to $\frac{1}{2} mv^2$; hence,

$$E - U = \frac{1}{2} mv^2,$$

and combination of this result with (99) leads to

$$\nabla^2 \psi + \frac{8\pi^2 m}{h^2} (E - U) \psi = 0 \quad (100)$$

This expression, generally known as the **Schrödinger wave equation**, has been used to solve a number of problems related to atomic and molecular structure.

Eigenvalues and Eigenfunctions.—In connection with the problem of standing waves in a stretched string, it is apparent that the solutions $f(x)$ of

* It should be noted that the actual velocity v of the particle is different from the velocity of propagation u of the hypothetical wave motion.

(96) can have significance only for certain definite values of the wave length λ ; these are the wave lengths corresponding to the normal modes of vibration of the string. The values are sometimes referred to as "proper values" or "characteristic values," but the hybrid term **eigenvalues** is most often used at the present time, especially in connection with the analogous quantities encountered in wave mechanics. The corresponding functions $f(x)$ which are solutions of (96) are called the **eigenfunctions** or **wave functions**. These functions must satisfy certain conditions which in the case of the stretched string are not difficult to understand. In the first place, $f(x)$ must be zero at each end of the string, since the system is fixed at these points and the vibration amplitude will be zero. In the second place $f(x)$ must be single valued and finite between the limits of x represented by the two ends of the string; in other words, at every point on the vibrating string the amplitude has a definite value at any given instant.

Just as the amplitude functions $f(x)$ in (96) have significance only for certain definite values of λ , so the functions ψ are satisfactory solutions of the wave equation (100) only for certain values, the eigenvalues, of the total energy E . For an atom these eigenvalues are the equivalent of the discrete set of energy values postulated by the Bohr theory. The occurrence of definite energy levels of the atom is thus a direct consequence of the wave mechanical treatment. The corresponding values of the function ψ are referred to as the wave functions or eigenfunctions of the Schrodinger equation. Just as with the functions $f(x)$ for the vibration of a stretched string, the eigenfunctions ψ of this equation must satisfy certain conditions: they must be single-valued, finite and continuous throughout the whole of the configuration space of the system under consideration, that is, for all possible values of the coordinates x , y and z , including infinity.

Physical Significance of the Wave Function.—In connection with the theory of the propagation of sound or light waves, the square of the amplitude is found to be proportional to the intensity of the sound or light, respectively. A somewhat analogous concept, modified by the requirements of the uncertainty principle, is made the basis of a physical interpretation of the wave function. The essential idea is that the probability of finding the system at a particular point in space is proportional to the square of the wave function, i.e., to ψ^2 .* Schrödinger originally suggested that for an electron the value of ψ^2 at any point could be interpreted as the electrical charge density at that point. According to this view the electron was to be regarded as "smeared out" into a cloud of electricity, the density of which would vary from place to place. M. Born (1926), however, proposed a statistical interpretation of the electron wave function, which has many advantages and has been widely adopted. The electron is still considered as a particle, and the quantity ψ^2 at any point is supposed to represent the probability of the electron being found at that point at a given instant. In a sense, therefore, the quantity ψ may be regarded as the wave mechanical equivalent of the electron orbit in the classical theory, and it is for this reason that the electron wave function is frequently referred to as an **orbital**, or as an **orbital wave function**. The importance of this interpretation is the admission that the wave picture and the corpuscular idea

* Strictly speaking this quantity should be the square of the absolute value of the wave function, i.e., $|\psi|^2$. However, most of the wave functions in atomic or molecular structure problems contain real terms only, and so ψ^2 may be employed.

are not mutually exclusive; an electron may be considered as a particle, while the wave equation determines the probability of finding it at any given point.

The Hydrogen Atom.—As applied to the hydrogen atom, and to ions, e.g., He^+ , Li^{++} and Be^{+++} , which are hydrogen-like, the Schrödinger wave equation can be reduced to the form

$$\nabla^2\psi + \frac{8\pi^2m}{h^2} \left(E + \frac{ze^2}{r} \right) \psi = 0, \quad (101)$$

where the potential energy term U in (100) has been replaced by $-ze^2/r$, as derived on page 46. The symbol m actually represents the reduced mass of the electron-nucleus system (p. 47), but this is approximately equal to the mass of the electron alone. The solution of (101) is complicated, but by introducing the condition that the eigenfunction ψ must be finite, single-valued and continuous, as stated above, the significant values of E , that is, the eigenvalues for the energy of the system, are found to be represented by

$$E_n = -\frac{2\pi^2ze^4m}{h^2n^2}, \quad (102)$$

which is identical with the Bohr equation (63). The value of n is determined by the facts that $n + l$ and $2l + 1$ must both be integral; hence, n and l must be integers (or l may be zero). Thus the two quantum numbers of the Bohr treatment are a direct consequence of the wave mechanical arguments. Further, an essential condition of the relationship between n and l is

$$l \leq n - 1, \quad (103)$$

so that l can have integral values from 0 to $n - 1$, inclusive. The azimuthal quantum number l , with values which include zero, as actually required by experiment (p. 63), is therefore derived theoretically. In this manner one of the serious difficulties of the Bohr-Sommerfeld atom is avoided.

By evaluating the wave functions corresponding to different values of the quantum numbers it is possible to obtain a picture of the probability distribution of the electron in a hydrogen-like atom. The fact that the probability has maxima and minima at various distances from the nucleus shows at once that definite electron orbits can have no reality in the wave mechanics. Nevertheless it is of interest to note that when l is large, for a given n , the electron distribution is mainly in one region, but it spreads out as l decreases, the maximum being at further distances from the nucleus. This corresponds, approximately, to the increasing ellipticity of the electron orbits in the Bohr-Sommerfeld theory with decreasing values of the azimuthal quantum number (p. 56).

Wave Mechanics and Spectral Lines.—Since orbits have no real existence in the atom of the new quantum mechanics, their place having been taken by stationary states or levels with definite energies, the interpretation of the emission of radiation as being due to the jump of an electron from one orbit to another cannot be applicable. The new mechanics, however, shows the Bohr postulate, that in a stationary state an atom does not radiate, to have a theoretical basis. The frequency of each spectral line may perhaps be regarded as a "beat" frequency between two states of the atom; this gives the same result as obtained by the Bohr treatment for a quantum transition between two orbits. It is not an easy matter to attribute a physical significance to the wave mechanical concepts, but the following crude pictorial representation may

be employed. If a quantity ψ^2 is assumed to represent the electrical charge density, then in any stationary state this remains constant; there is consequently no resultant movement of charge and no radiation of electromagnetic energy. When any state becomes excited by the absorption of energy, the function ψ is no longer constant but varies periodically, and consequently so also does the charge density which is determined by ψ^2 . The periodic variation in the distribution of electricity must be accompanied by the emission of radiation. It must now be supposed that when an atom absorbs energy, *two* of the possible stationary states are excited *simultaneously*; two sets of radiations are thus produced, and, being superimposed, give a beat vibration which has the frequency of the emitted radiation. If ν' and ν'' represent the frequencies of the two superimposed vibrations, then the frequency of the emitted beat radiation ν is given by

$$\nu = \nu' - \nu''. \quad (104)$$

The wave mechanical vibrations are related to the energies of the corresponding states by the usual quantum expressions $E' = h\nu'$ and $E'' = h\nu''$, so that

$$E' - E'' = h(\nu' - \nu'') = h\nu, \quad (105)$$

which is another of Bohr's postulates (p. 44).

Since wave mechanics leads to the same value for the energy in a stationary state as derived by Bohr, and also to the same connection between the energies of two combining states and the frequency of the emitted radiation, it clearly leads to the same equations for the frequencies of various spectral series which have been found to be in agreement with experiment. Further, the new mechanics gives a satisfactory interpretation of the principal and azimuthal quantum numbers, and it requires, as already noted, that values of zero should be possible for the latter. The other quantum numbers, required to explain the multiplicities of spectral terms and of the behavior of spectral lines in magnetic fields, also find a natural place. Apart from the possibility of zero values of l , the only other important change made by the new theories is that angular momenta are not simple multiples of $h/2\pi$ (see footnote, p. 62). If x represents the quantum numbers l , s or j , then the corresponding angular momentum is $\sqrt{x(x+1)}h/2\pi$, instead of $xh/2\pi$, as given by the older treatment. It appears, therefore, that all the conclusions reached from the Bohr theory which are in agreement with observation can be derived equally from wave mechanics by the use of rigid methods, although it must be admitted that the mathematical treatment is not simple. Finally, it may be mentioned that the selection rules, which give the restrictions for the change of quantum numbers, e.g., $\Delta l = \pm 1$ (see p. 60), follow directly from wave mechanical concepts, and so they are no longer of an empirical character. It should be realized, nevertheless, that wave mechanics, like the Bohr theory, has a postulational basis (p. 75); the essential difference is that the postulates in the former case are more consistent among themselves and with modern physical ideas than are those in the latter case. Further, they lead to more satisfactory conclusions.

The Atomic Spectrum of Helium.—The new quantum mechanics has not altered very seriously the results obtained from the Bohr-Sommerfeld treatment, and so the Bohr model may be retained for its pictorial value and for use in elementary calculations. In the interpretation of the spectrum of helium, however, the wave mechanical method represents an

appreciable quantitative advance. Incidentally it was in this connection that the concept of "resonance," which has no place in classical mechanics, was first introduced. This factor has played an important part in theories of valence and chemical structure, as will be evident later.

The spectrum of helium consists of two systems, almost completely independent, which were at one time believed to be produced by two different kinds of atoms. The terms of one system, the **parhelium terms**, are singlets, whereas the others, the **orthohelium terms**, are triplets. Since the helium atom has two electrons, each with a spin quantum number of $\pm\frac{1}{2}$, the resultant spin vector S will be 0 or 1; the multiplicities of the corresponding terms are $2S + 1$, and so it follows, according to the older theory (p. 68), that when the two electrons have spins of opposite sign (antiparallel) the terms will be singlets, and when they have spins in the same sense (parallel) the levels will be triplets. The ortho- and par-helium terms should, therefore, correspond to atoms in which the electron spins are parallel and antiparallel, respectively. This result is apparently satisfactory, but when an attempt is made to calculate the energies of the various levels the agreement with experiment is poor; the wave mechanical treatment, however, shows that a new factor must be considered.

Resonance.—In the normal (ground) state of the helium atom both electrons presumably have the same energy; in the formation of the spectrum an excited state must be involved, and in this condition the electrons must differ in energy. The wave function of one helium electron may be represented by $\psi(1)$ and that of the other by $\psi(2)$; if they are in different energy states they may be distinguished as $\psi'(1)$ and $\psi''(2)$, respectively, and the complete electron wave function for the excited helium atom may be written as their product, viz., $\psi'(1)\psi''(2)$. Suppose the electrons are interchanged so that electron (2) is now in the level indicated by one prime, and electron (1) is in the level distinguished by two primes; the wave function of the atom, which of course has the same energy as before, will then be $\psi(2)\psi''(1)$. The helium atom, therefore, has two possible electron wave functions for the same energy. Because of the properties of the Schrodinger wave equation, any linear combination of solutions of the equation must also be a solution. It follows, therefore, that two new wave functions, which are related to the sum and difference of the original functions, are possible; these are

$$\Psi_S = \frac{1}{\sqrt{2}} \{ \psi'(1)\psi''(2) + \psi(2)\psi''(1) \} \quad (106)$$

and

$$\Psi_A = \frac{1}{\sqrt{2}} \{ \psi'(1)\psi''(2) - \psi(2)\psi''(1) \}. \quad (107)$$

The function Ψ_S is said to be **symmetric**, since its sign remains the same when the coordinates of the electrons, designated by (1) and (2), are interchanged, whereas Ψ_A is **antisymmetric** because the sign would be reversed. The factor $1/\sqrt{2}$ is introduced for "normalization" purposes, in order to make the total probability of the existence of any electron equal to unity over the whole of space. The occurrence of the two new wave functions, given by (106) and

(107), is attributed to wave mechanical **resonance**. This result has a formal resemblance to the state of affairs existing when two strings vibrating, or two pendulums oscillating, with the same frequency are coupled together; the combined system then vibrates with two new frequencies, one being greater and the other less than the original value for the uncoupled systems. The two different possible wave functions Ψ_S and Ψ_A for the helium atom correspond to different energy values, which may be written

$$E_S = W + W_R \quad \text{and} \quad E_A = W - W_R, \quad (108)$$

where W is the normal energy, and W_R is called the **resonance energy**, or **exchange energy**. It appears, therefore, that two different forms of the excited helium atom should be possible, one with a somewhat greater energy than the other under similar conditions. From the spectrum of helium it is known that the ortho-levels have less energy than the par-levels with the same values of the quantum numbers n and l ; it follows, therefore, that the ortho-levels correspond to the antisymmetric wave function. Further considerations, based on the wave mechanical significance of the Pauli exclusion principle (see below), show that the symmetric levels must be singlets and the electron spins, or the equivalents of electron spin in wave mechanics, are antiparallel; the levels having antisymmetric wave functions must be triplets and the electron spins should be parallel. The rarity of intercombinations between the ortho- and par-helium states is to be attributed to the great difficulty of reversing electron spins.

Actual calculations, which are complicated, have been made of the various energies of the symmetric and antisymmetric levels with results in satisfactory agreement with those deduced from an analysis of the helium spectrum. This agreement is only possible as a result of the introduction of the resonance energy, as was first shown by W. Heisenberg (1926). It is in fact a fundamental generalization of wave mechanics that the resonance effect occurs whenever a particular system, involving either electrons or nuclei or both, can be represented by two or more wave functions corresponding to the same energy; different energy states are then possible, but they are not always stable as is the case with helium.²²

ELECTRONIC STRUCTURE OF ATOMS

The Pauli Exclusion Principle.—An interesting point concerning the helium spectrum is that there are no lines in the orthohelium spectrum corresponding to a transition involving the expected lowest term, namely 1^3S ; the lowest term actually concerned in the spectrum is 2^3S , representing a level in which the principal quantum number is 2. In seeking for an explanation of this observation, W. Pauli (1925) made a study of other spectra, and as a result enunciated the **exclusion principle**, known by his name; although it has yet no theoretical basis it has proved remarkably illuminating in connection with the problem of the arrangement of electrons in atoms, and that is the subject at which the discussion of spectra given here is really aimed. According to Pauli, *it is impossible for any two electrons in the same atom to have their four quantum numbers identical*; in the language of wave mechanics, the principle means that in an atomic system the *complete* wave function of every electron, that is the product of the orbital and spin functions, must be antisymmetric. It can be

readily shown that this limitation accounts for the nonexistence of the 1^3S state of helium. The atom of helium has two electrons, and if the principal quantum number (n) is 1 for both, l and m can only be 0 in each case, so that if the Pauli exclusion principle is to hold the spin quantum numbers of the two electrons must be opposite in sign. The electrons may consequently be described by their quantum numbers as follows:

- | | | | | |
|----|---------|---------|---------|----------------------|
| 1. | $n = 1$ | $l = 0$ | $m = 0$ | $s = +\frac{1}{2}$, |
| 2. | $n = 1$ | $l = 0$ | $m = 0$ | $s = -\frac{1}{2}$. |

Since the two electrons spins are antiparallel it follows that only the parhelium state is possible when n is 1. The orthohelium state, in which the electron spins are parallel, that is, both $+\frac{1}{2}$ or both $-\frac{1}{2}$, is clearly excluded by the Pauli principle if n is 1; when n is 2, however, one or both electrons being raised to a higher energy level, l can be 0 or 1, and m can be 0 or ± 1 , so that it is possible for two electrons to have parallel spins without violating the exclusion principle. The lowest possible orthohelium level must, therefore, have a principal quantum number of 2.

The combination of an antisymmetric orbital function with the wave function for parallel spins gives a complete wave function that is antisymmetric; so also does a symmetric orbital combined with an antiparallel spin function. Since the orbital function for orthohelium is antisymmetric, while that of parhelium is symmetric, the spin assignments given above are in harmony with the wave mechanical form of the Pauli principle.²³

Electron Groups.—It has been already inferred from a study of optical and X-ray spectra that the electrons in an atom arrange themselves in a series of groups or shells, and an application of the Pauli principle has thrown much light on the nature of this arrangement. The first comprehensive scheme of electronic architecture was devised by I. Langmuir (1919), who started with the fundamental idea, based on the earlier work of J. J. Thomson (1904), W. Kossel (1916), and G. N. Lewis (1916),* that the inert nature of the rare gases of the atmosphere, i.e., the zero group of the periodic table, was to be attributed to the possession by their atoms of completed electron shells. The number of electrons in an atom is equal to its atomic number (p. 25), and since the atomic numbers of the elements of the zero group are 2, 10, 18, 36, 54 and 86, these figures were taken to represent the completion of successive electron shells. The number of electrons in each shell should then be obtained by subtracting from the total in any inert gas atom the number in the preceding inert gas; this gives 2, 8, 8, 18, 18, 32 for the numbers of electrons in successive complete groups. These figures, of course, are also the numbers of elements in the successive periods of the periodic table, so that each represents the gradual filling up, from alkali metal to inert gas, of an electron shell. By means of these arguments Langmuir was able to account for the variation of chemical properties throughout the periodic

* The general idea was envisaged by Lewis in 1902, but not published until later.

system, and also for the presence of the transition elements of the eighth group, when the shell of 18 electrons succeeded the two groups of 8, and for the rare-earth elements, when the group of 32 electrons was developing. It was suggested that the first 18 electrons of this shell were arranged in such a manner as to imitate the two completed layers of 18 electrons immediately below, but that at the nineteenth electron a new type of arrangement was necessary; on this basis the element of atomic number $2 + 8 + 18 + 18 + 18 = 72$ should have been the last of the rare-earth series.

A modification of Langmuir's views was proposed by C. R. Bury (1921), who considered the maximum numbers of electrons in successive completed shells to be 2, 8, 18 and 32, but that the maximum number in the *outermost* shell of any atom is only 8; a shell was believed to contain more than this number of electrons only when there was a layer forming beyond it. According to Langmuir a new electron shell commenced only when the previous one was completely filled, but in the modified theory a new layer must form whenever the hitherto outermost shell attains 8 electrons. This means that the fourth shell commences to be formed after the third contains 8 electrons, although the latter is capable of holding 18. The gradual filling up of the third shell then takes place in the *fourth* period of the periodic system. Bury suggested that this would commence with titanium, the structure of which should be represented by the electronic arrangement 2, 8, 9, 3, whereas Langmuir had proposed 2, 8, 8, 4. The difference between the two points of view is best illustrated in connection with the elements of the rare-earth group. According to Bury, the electronic structure of lanthanum should be 2, 8, 18, 18, 9, 2, in which both fourth and fifth shells, as well as the sixth, are incomplete; but in the subsequent elements, those of the rare earths, the electrons enter the fourth shell until it attains its maximum of 32 electrons. The last of the rare-earth series should then have the structure 2, 8, 18, 32, 9, 2, and its atomic number should be 71, whereas Langmuir expected the element of atomic number 72 to be the last of the rare-earth group.

Almost simultaneously with the development of Bury's theory from chemical considerations, N. Bohr (1921) proposed a very similar scheme for the electronic structures of atoms based on a detailed study of their spectra. This scheme also required the element of atomic number 71 to be the last rare-earth element, while 72 should be a homologue of zirconium; experiment has shown that the expectations of Bohr and Bury are correct. D. Coster and G. Hevesy (1922) examined the characteristic X-rays obtained from zirconium minerals, and found in almost every instance lines of frequency corresponding to those to be expected, according to Moseley's law (p. 25), for an element of atomic number 72. This work led to the discovery of the new element "hafnium," which is definitely a homologue of zirconium and not a rare earth. This evidence favors the Bohr-Bury point of view, and supports the supposition that in many instances a new outer sheath of electrons can commence to form

while the inner ones are still incomplete. In the further development of his theory, Bohr suggested that each group was divided into subgroups capable of containing equal numbers of electrons, but alternative views were proposed by J. D. Main Smith (1924) and E. C. Stoner (1924), and the enunciation of the Pauli exclusion principle indicated that the latter were probably correct.²⁴

Electron Groups and Subgroups.—In the first electron shell, sometimes known as the *K*-shell (p. 70), the principal quantum number n is 1, so that for all electrons in this group both l and m must be 0.* Since the three quantum numbers are the same the Pauli principle indicates that the electrons in the *K*-shell must have different spin numbers. There are, however, only two possible values for s , viz., $+\frac{1}{2}$ and $-\frac{1}{2}$, so that only two electrons can occupy the first main quantum group. In the second group, or *L*-shell, where n is 2, l can be 1 or 0; in the former case m can be -1 , 0 or $+1$, but in the latter it can only be 0. Since there are two possible values of s for every combination of l and m , there are eight, and no more, different arrangements of the four quantum numbers. The maximum number of electrons in the second shell is thus eight, according to the requirements of the Pauli principle. The distribution of quantum numbers in the first two main groups is shown below:

FIRST QUANTUM GROUP.

K-shell: $n = 1$

l	0	0
m	0	0
s	$+\frac{1}{2}$	$-\frac{1}{2}$

SECOND QUANTUM GROUP.

L-shell: $n = 2$

l	0	0	1	1	1	1	1	1
m	0	0	-1	-1	0	0	+1	+1
s	$+\frac{1}{2}$	$-\frac{1}{2}$	$+\frac{1}{2}$	$-\frac{1}{2}$	$+\frac{1}{2}$	$-\frac{1}{2}$	$+\frac{1}{2}$	$-\frac{1}{2}$

In the *M*-shell n is 3, so that the possible values of l are 0, 1 and 2; if l is 0 then m must be 0; if l is 1, then m may be $+1$, 0 and -1 ; and if l is 2 then m may be $+2$, $+1$, 0, -1 and -2 . There are thus nine different combinations of l and m , and for each there can be two values of s , giving a maximum of 18 electrons in this shell. The detailed distribution is as follows:

THIRD QUANTUM GROUP.

M-shell: $n = 3$

l	0	0	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2
m	0	0	-1	-1	0	0	+1	+1	-2	-2	-1	-1	0	0	+1	+1	+2	+2
s	$+\frac{1}{2}$	$-\frac{1}{2}$	$+\frac{1}{2}$	$-\frac{1}{2}$	$+\frac{1}{2}$	$-\frac{1}{2}$	$+\frac{1}{2}$	$-\frac{1}{2}$	$+\frac{1}{2}$	$-\frac{1}{2}$	$+\frac{1}{2}$	$-\frac{1}{2}$	$+\frac{1}{2}$	$-\frac{1}{2}$	$+\frac{1}{2}$	$-\frac{1}{2}$	$+\frac{1}{2}$	$-\frac{1}{2}$

In exactly the same manner it may be shown that the fourth quantum group, or *N*-shell, can contain a maximum of 32 electrons; of these, two have $l = 0$, six have $l = 1$, ten have $l = 2$, exactly as in the third group, and in addition there can be fourteen for which l is 3. By the Pauli principle, therefore, the first four electron shells should be capable of accommodating a maximum of 2, 8, 18 and 32 electrons, respectively, in exact agreement with the suggestions made by Bury and Bohr.

* For the relation between the possible value of l and m to n , see pp. 58, 63.

Just as the main electron groups are determined by the principal quantum number, so the division into subgroups may be expressed by the l -values; it is immediately evident that the subgroups do not contain equal numbers of electrons. In order to describe the subgroups the following system of abbreviation is used. It has been shown previously (p. 69) that the term letters S, P, D, F, G , etc., refer to atomic states for which the values of L are 0, 1, 2, 3, 4, etc., respectively; the small letters, s, p, d, f, g , etc. are, therefore, employed to represent different electrons having the corresponding l -values. These letters are generally preceded by the principal quantum number, so as to give a more complete description; thus $3d$ represents an electron having $n = 3$ and $l = 2$. The symbol s is used for an electron for which l is 0, as well as for the spin quantum number, as seen above, but confusion is unlikely. If there are several electrons having the same value of n and l , that is to say, they are in the same subgroup, the number of such electrons is inserted as a superscript; for example, the group of six electrons for which n is 3 and l is 1 would be represented by the symbol $3p^6$. The first four quantum groups when complete would then be described in the following manner:

<i>K</i> -shell	$1s^2$,
<i>L</i> -shell	$2s^2, 2p^6$,
<i>M</i> -shell	$3s^2, 3p^6, 3d^{10}$,
<i>N</i> -shell	$4s^2, 4p^6, 4d^{10}, 4f^{14}$.

It is seen that successive subgroups are not equal, but contain 2, 6, 10 and 14 electrons, respectively. The same distribution applies to the higher levels, i.e., *O*-*P*- and *Q*-shells, but it will be evident that not more than the first three subgroups of these can be involved in the known elements.

Electronic Structures of the Elements.—With the data now available it is possible to consider the detailed electronic structures of the elements, use being made chiefly of optical and X-ray spectra, and of magnetic properties. In connection with the former it is important to find out from an analysis of the spectra, and their behavior in a magnetic field, what is the nature of the lowest term. This will represent the state of the atom with the least energy, and hence may be identified with the normal, or ground, state. Such information is often of considerable help in determining the distribution of electrons among the various subgroups. In order to elucidate their structural arrangement it may be imagined that the electrons, equal in number to the nuclear charge, are added one at a time; it is then required to find in which group each successive electron is accommodated. The complicated procedure involved can be greatly simplified by making the assumption, borne out by the regularity of X-ray spectra throughout the periodic system, with a few possible exceptions, that the main inner structure of any atom is the same as that of the preceding one; it is then usually only necessary to determine the position of the one additional electron which distinguishes the two elements. The essential principle adopted is based on the view that in

passing from one element to the next, the additional electron enters the particular subgroup in which it becomes most firmly bound, thus leading to the most stable structure.

First Period.—The hydrogen atom contains one electron only, and in the ground state it is probably in the lowest possible level for any electron, viz., $n = 1$, $l = 0$; that is, the only electron is $1s$, so that the lowest observed term, the ground term, should be $^2S_{1/2}$. For reasons already given (p. 60) hydrogen does not show the four series of lines associated with other elements, but the lowest observed spectroscopic term definitely corresponds to the $S_{1/2}$ terms of the alkali metals. The ground state of helium, which has two electrons, is represented by a 1S_0 term; this would agree with the configuration $1s^2$, that is, with both electrons having $n = 1$ and $l = 0$. The resultant azimuthal quantum number L is consequently 0, so that an S term is to be expected; further, since the Pauli principle requires the electrons to have antiparallel spins, the resultant spin and inner quantum numbers (S and J) must be zero (p. 67), so that the 1S_0 term is explained. It will be evident that the parhelium state is the ground state of helium.

Second Period.—Since the K -shell can hold only two electrons the L -shell must commence with the third element, lithium. The first two electrons are undoubtedly in the K -shell, since the first spark spectrum, attributed to Li^+ , is closely similar to that of the normal helium atom. The lowest observed level of lithium, as with the other alkali metals, is a $^2S_{1/2}$ term, and since only electrons in incomplete shells contribute to the resultant quantum numbers (see p. 67), it follows that the third electron has $n = 2$ and $l = 0$, that is to say, it is in a $2s$ state. The structure of the normal lithium atom is thus $1s^22s$. The spectrum of singly ionized beryllium, Be^+ , is analogous to that of lithium, so it may be supposed that the first three electrons of beryllium are similar to those of lithium, and the state of the fourth remains to be determined; this is probably also $2s$, leading to the configuration $1s^22s^2$. Since any s^2 group represents a complete subgroup, the lowest level for beryllium, like that of helium, should be a 1S_0 term, and this is in fact found to be the case. With boron it is necessary that the fifth electron should enter a new subgroup, and the $2p$ state appears most probable, giving the arrangement $1s^22s^22p$. As is to be expected, therefore, the lowest spectroscopic level is a $^2P_{1/2}$ term; since $l = 1$ and $s = \frac{1}{2}$ for the one p electron, the lowest j value is $\frac{1}{2}$. In the other elements of this period it appears that further electrons are added in the $2p$ level which is completed when neon, atomic number 10, is reached; its configuration is $1s^22s^22p^6$, and its ground term is 1S_0 , as for helium with its complete shell. Carbon is a specially interesting case to which the arrangement $1s^22s^22p^2$ is ascribed; the ground state is a 3P_0 term, so that the resultant spin quantum number must be 1. Since the $1s^2$ and $2s^2$ electrons contribute nothing it follows that the two $2p$ electrons must have parallel spins; reference to this point will be made later. The structures of the atoms of the first two periods may now be summarized as shown in Table 7.

TABLE 7. ELECTRONIC STRUCTURES IN THE FIRST TWO PERIODS

Atomic Number	Element	K-shell ($n = 1$)	L-shell ($n = 2$)		Ground Term
		s	s	p	
1	H	1			$^1S_{1/2}$
2	He	2			1S_0
3	Li	2	1		$^2S_{1/2}$
4	Be	2	2		1S_0
5	B	2	2	1	$^2P_{1/2}$
6	C	2	2	2	3P_0
7	N	2	2	3	$^4S_{3/2}$
8	O	2	2	4	3P_2
9	F	2	2	5	$^2P_{3/2}$
10	Ne	2	2	6	1S_0

Third Period.—In the third period the s and p levels of the M -shell, for which n is 3, are built up in a manner exactly analogous to that involved in the filling of the L -shell. With argon the $3s$ and $3p$ subgroups are complete thus giving a 1S_0 ground term, as for the other inert gases. The electronic configurations of the atoms of the third period, which present no novel features and are exactly analogous to those in the second period, are shown in Table 8; the abbreviation (Ne) is used to represent the neon structure, viz., $1s^22s^22p^6$.

TABLE 8. ELECTRONIC STRUCTURES IN THE THIRD PERIOD

Atomic Number	Element	K- and L-shells	M-shell ($n = 3$)		Ground Term
			s	p	
11	Na	(Ne)	1		$^2S_{1/2}$
12	Mg	(Ne)	2		1S_0
13	Al	(Ne)	2	1	$^2P_{1/2}$
14	Si	(Ne)	2	2	3P_0
15	P	(Ne)	2	3	$^4S_{3/2}$
16	S	(Ne)	2	4	3P_2
17	Cl	(Ne)	2	5	$^2P_{3/2}$
18	A	(Ne)	2	6	1S_0

Fourth Period.—In commencing the fourth period there are two possibilities to consider: in potassium, atomic number 19, the additional electron can either go into the $3d$ shell, which has yet no electrons in it, or else it can start a $4s$ level, leaving the $3d$ to be filled later. The ground term of potassium is, however, $^2S_{1/2}$ like that of sodium and lithium, and so there is no doubt that the last added electron must belong to the $4s$ group; the configuration of the atom is thus $1s^22s^22p^63s^23p^64s$. This conclusion implies that the $4s$ electron is more strongly bound, that is to say, it has a lower potential energy, than the $3d$ electron, in spite of its higher principal quantum number. In terms of the Bohr model, a $3d$ orbit, for which $n = 3$ and $l = 2$, is circular whereas the $4s$ should be a very eccentric ellipse; in the latter case the electron approaches very close

to the nucleus at times, so that the strong binding is quite comprehensible. Consideration of both arc and spark spectra of calcium, that is of Ca and Ca^+ , shows that in the normal atom the additional electron has also entered the $4s$ level, the spectrum of calcium being analogous to that of beryllium and magnesium.

With the next element, scandium, a marked change occurs: the ground term instead of being 2P , like aluminum, is actually 2D . Further, the lowest level in the spectrum of Sc^{++} , with 19 electrons, i.e., one in excess of argon, is a 2D term and not a $^2S_{1/2}$ term like normal potassium which it might have been expected to resemble. It is evident that the 19th electron must be in a $3d$ level; in other words, with scandium the third subgroup in the M -shell is beginning to fill up. From the spectra of singly ionized and normal scandium it appears that the 20th and 21st electrons are, however, in the $4s$ level, so that the electronic structure of the atom is $(\text{Ne})3s^23p^63d4s^2$, where (Ne) represents the arrangement of electrons as present in the neon atom. This is in harmony with the 2D ground term, since only the $3d$ electron contributes to the resultant angular momentum and hence L is 2 and S is $\frac{1}{2}$. The view that an inner level previously unoccupied, commences to fill up at scandium is also in agreement with the observation that at this element there is a change in the slope of the curve in which $\sqrt{\nu}/R$ for the M -series of X-rays is plotted against atomic number (see p. 72). It appears from the results with the Sc^{++} ion that, owing to the increase in the nuclear charge, the 19th electron is more firmly bound in the circular $3d$ orbit than it would be in the elliptical $4s$ orbit, but the shielding influence of this electron results in the 20th and 21st electrons being more firmly held in the $4s$ level than they would be if they also entered the $3d$ subgroup.

The spectra of Ti^{+++} and of V^{++++} are very similar to those of Sc^{++} , so that in these also the 19th electron is in a $3d$ state; of the three additional electrons in the neutral titanium atom, one is $3d$ and the other two are $4s$. In neutral vanadium there are three $3d$ and two $4s$ electrons. This type of configuration, involving the steady filling up of the $3d$ level, continues regularly up to nickel, atomic number 28, having the structure $(\text{A})3d^84s^2$, where (A) is the electronic arrangement as in argon. It appears, however, that chromium behaves exceptionally, for it has only one $4s$ electron (see Table 9); the reason for this conclusion is, in brief, as follows. It will be seen from Table 6 that the most prominent lines in the spectra of chromium are septets, and a study of the very complex spectra of neutral and ionized chromium has led to the conclusion that the ground state is a 7S term. Such a result implies that the resultant spin quantum number is 3, which is only possible if the six electrons in excess of the argon structure all have spins in the same direction; it is inevitable, therefore, that five of the electrons should be in the $3d$ subgroup and one in the $4s$ level, for in no other reasonable way could the six spins be in the same sense.

From spectroscopic and other considerations it is evident that after

nickel, that is at copper, there is another change in structural type. Instead of nine $3d$ and two $4s$ electrons, to follow the $3d^8 4s^2$ of nickel, there are now ten $3d$ electrons, thus completing this subgroup, and only one $4s$, thus (A) $3d^{10} 4s$. The lowest level of the spectrum of the neutral copper atom is thus a 2S term, like that of the alkali metals, since there is only one s electron apart from completed shells. Further, the cuprous ion Cu^+ is diamagnetic, showing that it consists of complete electron shells only.* It appears from the spectrum, however, that there is an excited 2D level of the neutral copper atom, represented by (A) $3d^9 4s^2$, which differs by a small amount of energy (0.15 e.v.) from the 2S ground level; the loss of the two $4s$ electrons results in the formation of the Cu^{++} ion which has a resultant magnetic moment and is consequently paramagnetic. The relative ease of obtaining the cupric ion in solution is to be attributed to the heat of hydration which supplies the necessary energy for exciting the copper atom from the 2S to the 2D state. From copper onward to krypton the $4s$ and $4p$ subgroups of the N -shell are gradually filled up in a normal manner, similar to that which occurs in the second and third periods (Tables 7 and 8). The complete structures of the elements of the fourth period are recorded in Table 9; the symbol (A) is used to represent the argon configuration, viz., $1s^2 2s^2 2p^6 3s^2 3p^6$.

TABLE 9. ELECTRONIC STRUCTURES IN THE FOURTH PERIOD

Atomic Number	Element	K-, L- and M-shells	M-shell ($n = 3$)	N-shell ($n = 4$)		Ground Term
			d	s	p	
19	K	(A)		1		$^2S_{1/2}$
20	Ca	(A)		2		1S_0
21	Sc	(A)	1	2		$^2D_{3/2}$
22	Ti	(A)	2	2		3F_2
23	V	(A)	3	2		$^4F_{3/2}$
24	Cr	(A)	5	1		7S_3
25	Mn	(A)	5	2		$^6S_{5/2}$
26	Fe	(A)	6	2		5D_4
27	Co	(A)	7	2		$^4F_{9/2}$
28	Ni	(A)	8	2		3F_4
29	Cu	(A)	10	1		$^2S_{1/2}$
30	Zn	(A)	10	2		1S_0
31	Ga	(A)	10	2	1	$^2P_{1/2}$
32	Ge	(A)	10	2	2	3P_0
33	As	(A)	10	2	3	$^4S_{3/2}$
34	Se	(A)	10	2	4	3P_2
35	Br	(A)	10	2	5	$^2P_{3/2}$
36	Kr	(A)	10	2	6	1S_0

It will be observed from Table 9 that the elements of atomic number 21 to 28, inclusive, form an exceptional group, there being a gradual filling up of the inner $3d$ group while the outer $4s$ level is unchanged. It has been already noted (p. 72) that some of the X-ray spectra of these "transitional" elements are abnormal; in addition, they form colored

* A discussion of para- and dia-magnetism and their relation to electronic structure will be found in Chapter VIII.

ions which are paramagnetic. In a sense, "cupric" copper, $(A)3d^94s^2$, may be regarded as belonging to the same series; the ion is colored when hydrated and has paramagnetic properties. The elements of the transitional group, apart from scandium, show variability of valence, probably because the strength of binding of the $3d$ electrons is very little different from those in the $4s$ level, so that excited states having little energy difference from the ground state, as with copper, are possible.

Fifth Period.—The first element of this period, rubidium, has exactly the same type of spectrum as the other alkali metals, and so it is evident that the additional electron enters the O -shell, i.e., $5s$, although the $4d$ and $4f$ levels are vacant. The structures of the first four elements of the fifth period are, in fact, parallel with those of the corresponding elements of the fourth period, for in strontium there are two $5s$ and no $4d$ electrons, but in the next element, yttrium, electrons commence to enter the inner ($4d$) group. The gradual filling up of this level continues up to palladium, although the structures, as deduced from the spectra, are not quite the same as for the corresponding elements in the fourth period. Nevertheless, the series of elements from yttrium to palladium, atomic numbers 39 to 46, forms a transition group having paramagnetic and colored ions of variable valence, similar to that in the fourth period. The frequencies of the characteristic X-rays do not vary in the normal manner (p. 72), and the results indicate the filling up of an inner quantum level. The structures of the elements between silver and xenon follow exactly those of the corresponding elements in the preceding period, and also of the elements in the second and third periods. The complete details are given in Table 10, where (Kr) indicates the arrangement of electrons as found in krypton, viz., $1s^22s^22p^63s^23p^63d^{10}4s^24p^6$.

TABLE 10. ELECTRONIC STRUCTURES IN THE FIFTH PERIOD

Atomic Number	Element	<i>K-, L-, M-</i> and <i>N</i> -shells	<i>N</i> -shell (<i>n</i> = 4) <i>d</i>	<i>O</i> -shell (<i>n</i> = 5) <i>s</i> <i>p</i>		Ground Term
37	Rb	(Kr)		1		$^2S_{1/2}$
38	Sr	(Kr)		2		1S_0
39	Y	(Kr)	1	2		$^2D_{3/2}$
40	Zr	(Kr)	2	2		3F_2
41	Nb	(Kr)	4	1		$^4D_{1/2}$
42	Mo	(Kr)	5	1		7S_3
43	Ma	(Kr)	6	1		$^6D_{5/2}$
44	Ru	(Kr)	7	1		5F_5
45	Rh	(Kr)	8	1		$^4F_{5/2}$
46	Pd	(Kr)	10			1S_0
47	Ag	(Kr)	10	1		$^2S_{1/2}$
48	Cd	(Kr)	10	2		1S_0
49	In	(Kr)	10	2	1	$^2P_{1/2}$
50	Sn	(Kr)	10	2	2	3P_0
51	Sb	(Kr)	10	2	3	$^4S_{3/2}$
52	Te	(Kr)	10	2	4	3P_2
53	I	(Kr)	10	2	5	$^2P_{3/2}$
54	Xe	(Kr)	10	2	6	1S_0

Since the $4d$ level tends to be filled up more readily than is the case with the corresponding $3d$ subgroup in the fourth period, as will be apparent by a comparison of the structures of nickel and its homologue palladium, for example, it is probable that the electrons are somewhat more tightly bound in the $4d$ than in the $5s$ level. It is not surprising to find, therefore, that silver is mainly univalent; a few bivalent compounds are known, however, and these are paramagnetic like those of bivalent copper (see p. 88).

Sixth Period.—The first three elements in the sixth period are again similar to those in the three previous periods. In spite of the fact that the $4f$, $5d$ and $5f$ subgroups are still vacant, the additional electron in cesium occupies the $6s$ level, and in the next element, barium, there are two $6s$ electrons. With lanthanum, atomic number 57, however, one electron enters the inner ($5d$) level, and frequencies of the O -series of X-rays commence to be abnormal; the general behavior is thus parallel with that observed in the fourth and fifth periods. Following lanthanum, that is at cerium, there begins the series of rare-earth elements, and in these it is very probable that the filling of the $4f$ subgroup, which according to the Pauli principle is capable of holding 14 electrons, takes place; the binding in this inner group is apparently stronger than in the still incomplete subgroups of the O - and P -shells. The spectra of the rare-earth elements are complex and have not been completely analyzed, but the properties of these substances suggest a steady increase in the $4f$ electrons while the outer groups, which mainly determine the chemical characteristics, remain unchanged. It will be seen that there should be 14 rare-earth elements, the last being lutecium, atomic number 71; with element number 72 a new type of structure and new properties must commence, since the N -shell is now complete. As previously recorded (p. 82) this actually occurs in the element hafnium which is definitely a homologue of zirconium as the suggested structure given below implies.

From hafnium to platinum the situation is somewhat analogous to that between titanium and nickel in the fourth period, and between zirconium and palladium in the fifth period; there is a gradual filling up of the inner $5d$ level. The structure of gold is like that of copper and silver, but as the binding in the $5d$ group does not appear to be particularly strong, the existence of a trivalent ion, involving the loss of three electrons, can be understood. From gold on to radon the $6s$ and $6p$ groups steadily fill up, with two and six electrons, respectively, as in the analogous elements of the previous periods. The complete electronic arrangements of the elements of the sixth period are summarized in Table 11, where (Pd) represents the configuration in palladium, viz., $1s^2 2s^2 2p^6 3s^2 3p^6 3d^{10} 4s^2 4p^6 4d^{10}$. It should be noted that (Pd) $5s^2 5p^6$ is the xenon structure. Attention may be called to the fact that there is some uncertainty concerning the ground terms of a number of the rare-earth elements.

An examination of the term symbols for the ground states of the

TABLE 11. ELECTRONIC STRUCTURES IN THE SIXTH PERIOD

Atomic Number	Ele- ment	<i>K</i> -, <i>L</i> -, <i>M</i> -, and <i>N</i> -shells	<i>N</i> -shell (<i>n</i> = 4) <i>f</i>	<i>O</i> -shell (<i>n</i> = 5) <i>s</i> <i>p</i> <i>d</i>			<i>P</i> -shell (<i>n</i> = 6) <i>s</i> <i>p</i>	Ground Term
55	Cs	(Pd)		2	6		1	$^2S_{1/2}$
56	Ba	(Pd)		2	6		2	1S_0
57	La	(Pd)		2	6	1	2	$^2D_{3/2}$
58	Ce	(Pd)	1	2	6	1	2	3H_4
59	Pr	(Pd)	2	2	6	1	2	$^4K_{11/2}$
60	Nd	(Pd)	3	2	6	1	2	3L_6
61	Il	(Pd)	4	2	6	1	2	$^4L_{11/2}$
62	Sm	(Pd)	5	2	6	1	2	7K_4
63	Eu	(Pd)	6	2	6	1	2	$^8H_{3/2}$
64	Gd	(Pd)	7	2	6	1	2	9D_2
65	Tb	(Pd)	8	2	6	1	2	$^8H_{17/2}$
66	Dy	(Pd)	9	2	6	1	2	$^7K_{10}$
67	Ho	(Pd)	10	2	6	1	2	$^6K_{19/2}$
68	Er	(Pd)	11	2	6	1	2	$^5L_{10}$
69	Tm	(Pd)	12	2	6	1	2	$^4K_{13/2}$
70	Yb	(Pd)	13	2	6	1	2	3H_6
71	Lu	(Pd)	14	2	6	1	2	$^2D_{3/2}$
72	Hf	(Pd)	14	2	6	2	2	3F_2
73	Ta	(Pd)	14	2	6	3	2	$^4F_{3/2}$
74	W	(Pd)	14	2	6	4	2	5D_0
75	Re	(Pd)	14	2	6	5	2	$^6S_{5/2}$
76	Os	(Pd)	14	2	6	6	2	5D_4
77	Ir	(Pd)	14	2	6	9		$^3D_{5/2}$
78	Pt	(Pd)	14	2	6	9	1	3D_3
79	Au	(Pd)	14	2	6	10	1	$^2S_{1/2}$
80	Hg	(Pd)	14	2	6	10	2	1S_0
81	Tl	(Pd)	14	2	6	10	2 1	$^2P_{1/2}$
82	Pb	(Pd)	14	2	6	10	2 2	3P_0
83	Bi	(Pd)	14	2	6	10	2 3	$^4S_{3/2}$
84	Po	(Pd)	14	2	6	10	2 4	3P_2
85	—	(Pd)	14	2	6	10	2 5	$^2P_{3/2}$
86	Rn	(Pd)	14	2	6	10	2 6	1S_0

rare-earth elements shows that for most of them the *L*, *S* and *J* values are exceptionally high. The resultant magnetic moment of the electrons in the atom would thus be expected to be large, and the element and its compounds should be highly paramagnetic. That this is the case has been proved without doubt, and the measured susceptibilities of the trivalent rare-earth ions are in good agreement with the values calculated by the equations on p. 65, the resultant quantum numbers *L*, *S* and *J* replacing *l*, *s* and *j*, respectively. Some of the experimental data for the magnetic moment per g. ion are compared with the values determined from (87) by F. Hund (1925) and those which include Van Vleck's correction (Table 12). Except for the samarium and europium ions the two calculated results are almost identical, but in these instances the additional term is necessary to harmonize the experimental and theoretical magnetic moments.

TABLE 12. CALCULATED AND OBSERVED MAGNETIC MOMENTS OF RARE-EARTH IONS IN BOHR MAGNETONS

Ion	Ground Term	Magnetic Moment		
		Calculated		Observed
		Hund	Van Vleck	
Ce ⁺⁺⁺	$^3F_{5/2}$	2.54	2.56	2.5
Pr ⁺⁺⁺	3H_4	3.58	3.62	3.6
Sm ⁺⁺⁺	$^6H_{5/2}$	0.84	1.65	1.6
Eu ⁺⁺⁺	7F_0	0	3.51	3.5
Gd ⁺⁺⁺	$^8S_{7/2}$	7.94	7.94	7.9
Ho ⁺⁺⁺	5I_8	10.6	10.6	10.6
Yb ⁺⁺⁺	$^3F_{7/2}$	4.54	4.54	4.5
Lu ⁺⁺⁺	1S_0	0	0	0

The case of the lutecium ion is of interest, for although it is a true rare-earth ion all its electrons are in closed shells, viz., (Pd) $4f^{14}5s^25p^6$, so that it cannot be paramagnetic.

Seventh Period.—Once again the first three elements appear to be built up as in the earlier groups, an inner shell commencing to be filled at the third, i.e., actinium. A group analogous to the rare-earth elements might be expected to develop at this point in the seventh period, just as in the previous one, and the properties of the newly discovered elements 93, 94, 95 and 96 (p. 173) indicate that this is the case. The spectra of the radioactive elements which make up this period have not been analyzed and so the electronic structures are not known with certainty, but chemical considerations in general lead to the suggestions found in Table 13; (Rn) is the radon configuration.

TABLE 13. ELECTRONIC STRUCTURES IN THE SEVENTH PERIOD

Atomic Number	Element	K-, L-, M- N- and O-shells	P-shell ($n = 6$) d	Q-shell ($n = 7$) s	Ground Term
87	—	(Rn)		1	($^2S_{1/2}$)
88	Ra	(Rn)		2	(1S_0)
89	Ac	(Rn)	1	2	($^2D_{3/2}$)
90	Th	(Rn)	2	2	(3F_2)
91	Pa	(Rn)	3	2	($^4F_{3/2}$)
92	U	(Rn)	4	2	(5D_0)

NUCLEAR SPIN

Hyperfine Structure.—In addition to the fine structure of spectral lines due to multiplets, in many instances a **hyperfine structure** has been detected. To some extent it may be due to isotopes (cf. p. 147), but this does not account for (i) the presence of hyperfine structure in the spectra of heavy elements, where the isotope effect is very small; (ii) the fact that the structures of different lines are not always the same, as they should be if resulting from the presence of isotopes; and (iii) the existence of hyperfine structure with certain elements, e.g., bismuth, which have no stable isotopes. It is now generally accepted that hyperfine structure of this type can be explained by supposing the atomic

nuclei to have angular momenta equivalent to spin (W. Pauli, 1927); it is equal to $\sqrt{i(i+1)}h/2\pi$ units, where i , the nuclear spin number, is zero or an odd or even multiple of $\frac{1}{2}$. This spin momentum, through its associated magnetic moment, interacts with the resultant quantum number J of the electronic system, and produces, except when $i = 0$, for each value of J a series of values of total angular momentum F of the whole atom; thus,

$$F = J + i, \quad J + i - 1, \quad J + i - 2, \quad \dots, \quad J - i,$$

which accounts for the hyperfine structure. From a study of the spectral lines it has been possible to ascertain the spin quantum numbers of the nuclei of several atoms, and these have been confirmed, and others determined, by a study of the Zeeman effect of the hyperfine structure (cf. p. 63).

Spin Quantum Numbers.—In view of the relatively large mass of the nucleus; the magnetic moment associated with nuclear spin is very small and for most purposes it is negligible in comparison with that due to the angular momentum of the electrons (p. 67). In a strong inhomogeneous magnetic field, such as is used in the Stern-Gerlach experiment (p. 64), therefore, there is no coupling between the nuclear and electron angular momenta, and the splitting of the atomic ray is determined by the J values only. If one of the components, obtained in this way, is separated from the others and sent through a second field which is weak but extremely inhomogeneous, the i and J values are no longer uncoupled and a further splitting occurs into $2i + 1$ components. This principle has been used by I. I. Rabi and his collaborators (1933 *et seq.*) to determine nuclear spin quantum numbers.

Another method for the study of nuclear spin depends on an examination of the intensities of the lines in the band spectra of molecules (p. 572). Molecules, like atoms, possess electronic energy levels, which are subdivided into vibrational levels, representing different energies of vibration, and these are further divided into rotational energy levels. It is the latter which are responsible for the fine lines observed in the spectra of molecules. From theoretical considerations it has been shown that at normal temperatures the relative intensities of the rotational lines are determined by the nuclear spins of the atoms. If the two nuclei in a diatomic molecule are different, then each rotational level has a multiplicity equal to $(2i + 1)(2i' + 1)$, where i and i' are the respective spins; all the rotational lines have, therefore, equal intensities. For a molecule with two *identical* nuclei, of spin i , however, alternate levels have the multiplicities $(i + 1)(2i + 1)$ and $i(2i + 1)$, respectively; whether it is the odd or even rotational levels that have the respective multiplicities depends on a variety of factors which need not be considered here. The intensities of alternate rotational lines will consequently be in the ratio of $i + 1$ to i ; it is thus possible to determine the nuclear spin directly. In the spectrum of molecular hydrogen, for example, the intensities of alternate lines are in the ratio of 3 to 1; that is,

$$\frac{i + 1}{i} = \frac{3}{1}; \quad \therefore i = \frac{1}{2},$$

so that the spin of a hydrogen nucleus is $\frac{1}{2}$ unit. With deuterium, on the other hand, the alternate lines have intensities of 2 to 1, so that the nucleus has a spin of 1 unit. When the nucleus has no spin momentum, i.e., $i = 0$, alternate rotational levels must have a multiplicity of zero; that is to say, alternate rotational

lines are missing. This is the case with the molecules of oxygen in which both atoms have atomic masses of 16; the nuclei of these atoms have no nuclear spin. In the $^{16}\text{O}^{18}\text{O}$ molecules, however, the two nuclei are different, so that all the rotational lines must be present at equal intensity; this is in agreement with experiment.

The different methods described above give the same results and the spin numbers of a few elements are recorded in Table 14. As a general rule, to which deuterium and nitrogen are exceptions, elements of even atomic weight have zero nuclear spin.

TABLE 14. NUCLEAR SPIN QUANTUM NUMBERS

Atom	Mass Number	Nuclear Spin	Atom	Mass Number	Nuclear Spin
H	1	$\frac{1}{2}$	O	16	0
D	2	1	F	19	$\frac{1}{2}$
He	4	0	Na	23	$\frac{3}{2}$
C	12	0	Cl	35	$\frac{3}{2}$
N	14	1	Cl	37	$\frac{3}{2}$

Ortho- and Para-States.—A remarkable consequence of nuclear spin is the existence of **ortho-** and **para-states** of molecules with two identical nuclei; in the ortho-states the two nuclei have parallel spins, i.e., in the same direction, whereas in the para-states the nuclear spins are anti-parallel.* Nuclear spins, like those of electrons, are only reversed with difficulty, and consequently when they interact with radiation to produce spectra the ortho- and para-molecules each give rise to their own rotational levels. In fact the alternate lines are produced by ortho- and para-states, and the regular alternation of intensity is due to the different amounts of the two forms present at equilibrium. If a molecule has two nuclei which are not identical, or if the nuclei are identical but have no spin, then ortho- and para-states cannot exist.

A study of the spectrum of molecular hydrogen under ordinary conditions has shown that the more intense lines, which are produced by ortho-molecules, correspond to odd rotational levels in the ground state, whereas the less intense lines, due to the para-molecules, represent even rotational levels. From the intensities of the lines, therefore, it may be supposed that the normal gas consists of three parts of orthohydrogen to one of parahydrogen. As the temperature of the gas is lowered, and the energy content of the molecules decreases, the higher rotational levels, i.e., with large amounts of rotational energy, become less common, and the molecules tend to occupy the lower levels. Near the absolute zero, therefore, the majority of the gas molecules should occupy the zero rotational level. For present purposes zero is an even number, and so the molecules of hydrogen should be mainly in the para-state at low temperature, whereas at ordinary temperatures the gas contains only one-quarter of para-

* The prefixes "ortho" and "para," (Greek: ortho, *straight*; para, *side by side*), which are not really satisfactory, have been given by analogy with helium which exists in ortho- and para-states (p. 79) because of the different coupling of *electron* spins.

molecules. Owing to the difficulty of reversing nuclear spins the establishment of the low temperature equilibrium takes a considerable time, but if such time were allowed a change should be observed in the properties of hydrogen. Acting on the suggestion of E. U. Condon, hydrogen was kept at -188°C . for 197 days (W. F. Giauque and H. L. Johnston, 1928) and a small but definite change was noted in the triple-point pressure (cf. p. 465). In 1929, K. Bonhoeffer and P. Harteck discovered that equilibrium could be accelerated by the use of charcoal as catalyst. Hydrogen was adsorbed on charcoal and kept at the temperature of liquid hydrogen for a short time; the gas pumped off had a different heat conductivity, and hence a different heat capacity, from that of normal hydrogen. Almost simultaneously A. Eucken and K. Hiller (1929) found, by direct measurement, a change in the heat capacity of hydrogen that had been kept at -183°C . for some weeks.

These results definitely confirm the view that hydrogen molecules exist in two forms, and by modification of the catalytic method described above almost pure parahydrogen can be obtained. It can be stored in glass vessels for several weeks at ordinary temperatures without appreciable change, because of the difficulty of reversing the nuclear spin which would be necessary for the equilibrium proportion of 75 per cent of orthohydrogen to be formed. In the presence of oxygen or water as catalyst, or if the gas is heated to 800°C ., or an electric discharge passed through it, equilibrium is attained in a relatively short time. It will be seen shortly that although hydrogen should be entirely in the para-form at low temperatures, the maximum possible equilibrium proportion of the ortho-state is 75 per cent; consequently no enrichment of this form of hydrogen over the proportion present in the normal gas has been observed.

The physical properties of parahydrogen, e.g., vapor pressure, boiling point, and specific heat, differ from those of normal hydrogen; in the spectrum of the former, the alternate, more intense, rotational lines, which are due to the ortho-form, are of course missing. Further, since the nuclear spins in parahydrogen are antiparallel and so cancel each other, the nuclei do not contribute to the magnetic moment: in orthohydrogen, however, they make a definite contribution. The difference in magnetic properties is utilized to expedite the attainment of ortho-para equilibrium; paramagnetic substances, e.g., oxygen, sesquioxides of chromium and the rare-earth elements, and free radicals accelerate the change from ortho- to para-form, and *vice versa*.

Ratio of Para- and Ortho-Forms.—If N_0 is the number of diatomic molecules in the lowest rotational energy level, then by the Maxwell-Boltzmann distribution law (p. 272) the number in any other level is $N_0 e^{-\epsilon/kT}$ where ϵ is the rotational energy in the given level in excess of the lowest level (cf. p. 32). Each rotational level consists of a number of states, the energies of which are almost identical; for purposes of calculation such states may be grouped together and a **statistical weight**, representing the number of multiplets included, is applied to the level as a whole. If p_J is the statistical weight in any given rota-

tional level of quantum number J ,* then the actual number of molecules in this state is $p_J N_0 e^{-\epsilon_J/kT}$, where ϵ_J is the rotational energy in that level. The statistical weight of each rotational level is given by the product of the rotational multiplicity $(2J + 1)$ and nuclear spin multiplicity (p. 92). For hydrogen the latter quantity is $(i + 1)(2i + 1)$ for the ortho-molecules, i.e., with the odd rotational levels, whereas in the even levels, i.e., para-molecules, the spin multiplicity is $i(2i + 1)$. The ratio of the number of molecules in the para-state to those in the ortho-state is then

$$\frac{[\text{para-H}_2]}{[\text{ortho-H}_2]} = \frac{N_0 \sum_{J=0,2,4,6,\dots} p_J e^{-\epsilon_J/kT}}{N_0 \sum_{J=1,3,5,7,\dots} p_J e^{-\epsilon_J/kT}} = \frac{\sum_{J=0,2,4,6,\dots} i(2i + 1)(2J + 1)e^{-\epsilon_J/kT}}{\sum_{J=1,3,5,7,\dots} (i + 1)(2i + 1)(2J + 1)e^{-\epsilon_J/kT}}. \quad (109)$$

Since i is $\frac{1}{2}$ for hydrogen, it follows that

$$\frac{[\text{para-H}_2]}{[\text{ortho-H}_2]} = \frac{1 + 5e^{-\epsilon_2/kT} + 9e^{-\epsilon_4/kT} + 13e^{-\epsilon_6/kT} + \dots}{3(3e^{-\epsilon_1/kT} + 7e^{-\epsilon_3/kT} + 11e^{-\epsilon_5/kT} + \dots)}. \quad (110)$$

This equation can be evaluated by inserting the rotational energies in the various levels as determined from the spectrum of molecular hydrogen; the terms in both numerator and denominator are summed as long as they contribute appreciably to the total at different temperatures. Actually $e^{-\epsilon_J/kT}$ falls off rapidly as J increases, especially at fairly low temperatures, and so the number of terms to be included is not excessive; at 300° Abs. , for example, only the first five rotational levels of hydrogen need be taken into consideration. The estimation of the ratio of para- to ortho-molecules at various temperatures may be made in another manner which is somewhat less accurate but much simpler. The rotational energy of a rigid diatomic molecule, that is, one having a constant moment of inertia, can be expressed by the relationship (see p. 566) $\epsilon_J = J(J + 1)\hbar^2/8\pi^2 I$. For hydrogen, the moment of inertia I is $0.467 \times 10^{-40} \text{ g. cm.}^2$, and so the ϵ_J values for a series of integral values of J can be determined and hence (110) evaluated. The results for the amounts of ortho- and para-hydrogen at various temperatures, obtained by the accurate method, based on measurements of band spectra, are given in Table 15. It will be

TABLE 15. ORTHO-PARA COMPOSITION OF EQUILIBRIUM HYDROGEN

Temp.	Para-H ₂	Ortho-H ₂	Temp.	Para-H ₂	Ortho-H ₂
20° Abs.	99.82%	0.18%	100° Abs.	38.51%	61.49%
40°	88.61	11.39	120°	32.87	67.13
50°	76.89	23.11	170°	27.99	72.01
60°	65.39	34.61	230°	25.42	74.58
70°	55.83	44.17	273°	25.13	74.87
80°	48.39	51.61	∞	25.00	75.00

noted that as the temperature is raised the proportion of para-form in equilibrium hydrogen decreases from 100 per cent to a limiting value of 25 per cent; hence at ordinary temperatures hydrogen should contain three parts of

* The rotational quantum number J of a molecule should not be confused with the same symbol which gives the resultant angular momentum of an atom.

ortho- to one part of the para-form, in agreement with the intensities of alternate lines in the molecular spectrum. After about 230° Abs., it is seen that the ortho-para ratio changes very little with increasing temperature; the reason for this is that as the temperature increases the sum of the $(2J + 1)e^{-J(J+1)h^2/8\pi^2 I kT}$ terms for odd values of J becomes almost equal to the sum for even values. The ratio of numerator to denominator in (110) is then determined only by the ratio of the nuclear spin factors, i.e., 1 to 3.

Like hydrogen, its isotope deuterium exists in two forms; for this gas, however, the even rotational levels constitute the ortho-state, and the odd levels the para-form. At low temperatures deuterium consequently exists almost entirely in the ortho-state, whereas on raising the temperature a limiting ratio of para- to ortho-forms equal to $i/(i + 1)$, i.e., 1 to 2, since $i = 1$, is attained.

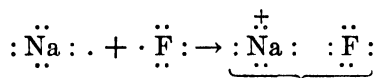
A temperature of about 230° Abs. is required before the limiting ortho-para ratio is reached in hydrogen, but 100° Abs. is sufficient for deuterium. The smaller the value of the term $h^2/8\pi^2 I$, which determines the magnitude of the rotational energy, as seen above, the lower the temperature at which the sum of the odd terms becomes equal to the sum of the even terms in the equation for the ortho-para ratio; that is, the larger the moment of inertia I the lower the temperature at which the limiting equilibrium ratio is attained. The physical significance of this fact is that when the moment of inertia is small the quantum of rotational energy, determined by $h^2/8\pi^2 I$, is large, and a fairly high temperature is required before the molecules occupy a reasonable number of rotational levels; if the rotational quantum is small, that is, I is large, however, the molecules will be spread over several levels at a relatively low temperature.

Since nitrogen and chlorine have nuclear spins (Table 14), the molecules $^{14}\text{N}_2$ and $^{35}\text{Cl}_2$ should exist in ortho- and para-states; that they actually do so is shown by the alternating intensities of the rotational lines in the molecular spectra. Owing to the relatively high moments of inertia of these molecules, the temperatures at which the limiting ortho-para ratios are attained are so low that enrichment, such as has been achieved with hydrogen and deuterium, is impossible because the substances are then solid. Since rotation in the solid state is not common, there is little hope of obtaining evidence for a change in the ortho-para ratio. It may be mentioned that symmetrical polyatomic molecules, such as water, deuterium oxide, cyanogen and acetylene, exist in ortho- and para-forms; they behave in a manner similar to hydrogen, deuterium and nitrogen molecules, since the other atoms, viz., carbon and oxygen, have no nuclear spins.²⁵

THE ELECTRONIC THEORY OF VALENCE

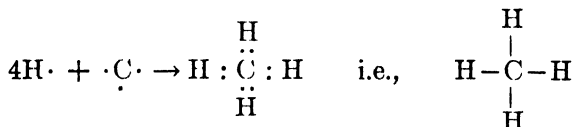
Electrovalence and Covalence.—Atomic spectra support the view that the inert gas type of electronic arrangement is one of exceptional stability (p. 81), and it is reasonable to assume, therefore, that in chemical combination atoms tend to acquire the inert gas structure (W. Kossel, 1916). The alkali metals, for example, have one electron in excess, and hence in combination they will lose this electron if possible; the tendency for these elements to form ions with a single positive charge is thus

readily understood. In an analogous manner it is possible to explain the formation of bivalent positive ions by the alkaline-earth metals, the atoms of which contain two electrons more than the corresponding inert gases. On the other hand, the halogens have one electron less than required for stability; the atoms will consequently attempt to gain an extra electron in chemical combination, thus forming singly charged negative ions. Similarly, oxygen and its homologues should tend to form bivalent negative ions. In the combination of an electropositive with an electronegative element, e.g., sodium and fluorine, the former is prepared to yield an electron to the latter, so that a positive sodium ion and a negative fluorine ion are formed: thus,*



In the crystal of sodium fluoride the two ions would be held together by electrostatic forces, and it is supposed that the same type of union, called **electrovalence** (I. Langmuir, 1919), exists in all simple salts. The evidence from crystal structure and from the study of electrolytes supports this view.

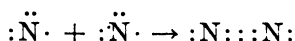
The above theory accounts satisfactorily for the formation of salt-like compounds by elements which are close to the inert gases in the periodic table, but with others further away difficulties arise. It is known from spectroscopic measurements, for example, that the ionizing potentials (p. 50) of the alkali and alkaline-earth elements are relatively low, but the energy required to remove three electrons from an element in the third group of the periodic system, e.g., boron, is very large, and it is improbable that a B^{+++} ion will form in chemical combination. Further, it is doubtful whether the nitrogen atom has sufficient electron affinity to take up three electrons and form the N^{---} ion. It was suggested by G. N. Lewis (1916) that in addition to the type of behavior described by Kossel, which applies to ionized compounds, the inert gas structure could be attained in chemical combination by the *sharing* of electrons in pairs. For example, in the formation of methane the four hydrogen atoms each contribute one electron and the carbon atom four electrons towards the formation of four pairs; thus,



In this way the carbon atom acquires the neon structure and the hydrogen atoms each have two electrons as in helium. This type of union, which

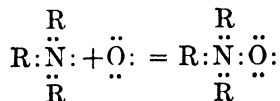
* The symbol of an element is used to represent the nucleus of an atom and its inner electron shells; thus F represents the fluorine nucleus and two 1s electrons. The dots represent electrons in the outer shell or shells.

is very common in organic compounds, clearly does not lead to the formation of ions, and to distinguish it from the type described previously it has been called **covalence** (I. Langmuir, 1919). There are two essential points to notice: firstly, that sharing of electrons occurs in pairs, each pair or **duplet** being equivalent to one conventional chemical bond, as shown by the formula of methane given above. Lewis was led to this suggestion by the striking fact that with but few exceptions, e.g., nitric oxide, nitrogen dioxide, chlorine dioxide and free radicals, all known compounds contain an even number of electrons. The second point is that there is a tendency in covalent union for each atom to acquire an outermost layer of eight electrons, i.e., an **electron octet**; this corresponds to the electronic arrangement in the inert gases. Hydrogen is, of course, an exception since to acquire the structure of helium only two electrons are required. It is by sharing electrons that two elements, each of which has less than eight electrons in its outer sheath, can both acquire complete octets; a nitrogen atom has only five outer electrons, but if two such atoms share three pairs between them then each has virtually a complete octet, viz.,



A single bond implies the sharing of one duplet, a double bond two duplets, and a triple bond three duplets, as in the nitrogen molecule. A quadruple bond would require the sharing of four pairs of electrons between two atoms; this is impossible unless the atom is flat, and so the nonexistence of such bonds is understood. It may be mentioned that although the octet rule is a useful general guide, it is probably not strictly applicable in all cases, for it appears that in certain instances groups containing more than eight electrons are possible. For example, in SF_6 the six fluorine atoms share electron pairs with the sulfur, and so the latter has an external group of twelve electrons; similarly, in PCl_5 the phosphorus atom has presumably ten electrons * in its outer shell.²⁶

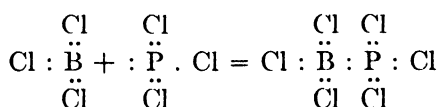
The Semipolar or Dative Bond.—For certain purposes it is convenient to distinguish a type of covalent link in which both electrons forming the duplet are supplied by one atom. For example, in the formation of amine oxides between a tertiary amine and oxygen, the pair of electrons for the bond between N and O comes entirely from the nitrogen atom; thus,



Since the R_3N molecule and the oxygen atom are originally electrically neutral, the transfer of an electron pair as shown will result in a tendency for the R_3N portion of the amine oxide to have a positive charge and the

* Crystals of phosphorus pentachloride apparently consist of PCl_4^- and PCl_4^+ units; in the former there are evidently twelve electrons in the outer shell of the phosphorus atom.

oxygen atom a negative charge. In other words, an electrical dipole will be produced which, in a sense, is equivalent to two ions held together by a covalent bond. For this reason the linkage is sometimes called a **semi-polar bond**, and the presence of the charges may be indicated by writing the formula in the following manner, $R_3\overset{+}{N}-\bar{O}$. Another name, preferred by some writers, is **dative bond** or **dative covalence**: the adjective "dative" refers to the fact that one atom "donates" or gives a pair of electrons towards the formation of the single bond. The terms **donor** and **acceptor** are often employed to describe the two atoms, or the molecules in which they are present, involved in a dative bond. Another name which has been proposed is **coordinate link**, for this type of bond is found in what are called coordination compounds. In the metal-ammines, for example, each of the nitrogen atoms of the ammonia molecules donates a pair of electrons to the central atom. The "coordination" or "auxiliary" valence of Werner is thus a form of covalent linkage. His "principal valence" really represents electrovalence (cf. p. 110). The term coordinate bond or link is best restricted to cases in which combination occurs, by means of a dative bond, between two molecules which can exist alone; this is true for most of the Werner coordination compounds. Similarly, many "molecular compounds" may be regarded as being formed in the same way. In its simple tervalent compounds the boron shell has only six electrons, and such substances readily form molecular compounds with donor molecules, e.g., NH_3 , PCl_3 and $(C_2H_5)_2O$. The nitrogen, phosphorus or oxygen atom, respectively, has a **lone pair** of electrons which it can donate to the boron atom, so that the latter may complete its octet. The formation of the compound between boron and phosphorus trichlorides may be represented by



The electron octets of the chlorine atoms are omitted so as to avoid confusion. In these cases an arrow is often used in the formula to indicate the direction in which the electron pair has been transferred, viz., $\text{Cl}_3\text{B} \leftarrow \text{PCl}_3$. In a sense the formation of complex halide ions, e.g., the borofluoride ion BF_4^- , may be regarded as involving a coordinate link between a BF_3 molecule and a fluorine ion, thus $\text{F}_3\text{B} \leftarrow \text{F}^-$. In the final ion there is no difference between the four fluorines, and so the negative charge may be regarded as centered on the boron atom, or perhaps as belonging to the molecule as a whole. This point may be seen by inserting the positive and negative signs associated with a semipolar bond, as shown above. The formation of the BF_4^- ion may then be represented by $\text{F}_3\text{B}-[\text{F}^{+-}]$, which is obviously equivalent to $\text{B}-\text{F}_4$ or $(\text{BF}_4)^-$. The formation of the ammonium ion from an ammonia molecule and a proton, H^+ , may be considered from the same point of view. In this case

the nitrogen acts as the donor of an electron pair to the proton; thus $\text{H}_3\text{N} \rightarrow \text{H}^+$, or $\text{H}_3\text{N}^+ - [\text{H}^-]$, which is N^+H_4 or $(\text{NH}_4)^+$.

Wave Mechanics and Valence.—Two main methods of the application of wave mechanics to electrons involved in valence bond formation have been developed, with interesting results in each case. In this connection an atomic electron wave function, or orbital (cf. p. 76), generally called an **atomic orbital**, is defined as the wave function for an electron moving in the field of only one atom. In the **homopolar method** of W. Heitler and F. London (1927), developed by L. Pauling (1931) and J. C. Slater (1931), it is assumed that in the formation of a molecule the orbital part of the wave function is a linear combination of the products of the separate atomic orbitals. Consider, for example, a diatomic molecule AB with two electrons which are of importance in the union of the atoms A and B; then the orbital wave function Ψ of the molecule may be represented by

$$\Psi = C\{\psi_A(1)\psi_B(2) + \psi_A(2)\psi_B(1)\}, \quad (111)$$

where ψ_A and ψ_B are the atomic orbitals of the electrons designated by (1) and (2), when in the field of the atoms A and B, respectively; the factor C is introduced for normalization purposes (p. 79). Since the two electrons are interchangeable (p. 79) there are actually two values for the molecular wave function, but for the present one only need be considered. In the second mode of treatment, known as the method of **molecular orbitals** (F. Hund, 1928; J. E. Lennard-Jones, 1929; R. S. Mulliken, 1931 *et seq.*), it is supposed that in the formation of a molecule the atomic orbital becomes a molecular orbital; this is defined as the wave function of an electron moving in the field of all the other electrons and the nuclei constituting the molecule, and generally expressed as the product of a linear combination of atomic orbitals. For the substance AB considered above, the molecular orbitals of the two electrons are thus assumed to be $a\psi_A(1) + b\psi_B(1)$ and $a\psi_A(2) + b\psi_B(2)$, where ψ_A and ψ_B are, as before, the atomic orbitals and a and b are coefficients determining the extents to which they take part in the molecular orbital. The wave function of the molecule may then be expressed by the product of the separate orbitals; thus,

$$\begin{aligned} \Psi &= \{a\psi_A(1) + b\psi_B(1)\} \{a\psi_A(2) + b\psi_B(2)\} \\ &= a^2\{\psi_A(1)\psi_A(2)\} + b^2\{\psi_B(1)\psi_B(2)\} + ab\{\psi_A(1)\psi_B(2) + \psi_A(2)\psi_B(1)\}. \end{aligned} \quad (112)$$

The third term of (112) is seen to be virtually identical with (111), leaving two additional terms which arise in the molecular orbital treatment; in each of these terms, however, the two electrons are represented as being associated with the same atom, viz., A in the first term and B in the second. The two extra terms in (112), therefore, give the wave functions for the molecule in the ionic states A^-B^+ and A^+B^- , respectively. They are consequently called **ionic terms**, and it is the presence of these terms which constitutes the essential difference between the two methods of defining the wave function of a molecule. The name homopolar method for the Heitler-London treatment originated in the fact that ionic terms are not involved; this is probably a weakness, for there is little doubt that such terms should frequently be included to some extent. On the other hand, the molecular orbital treatment often lays too much emphasis on the ionic contributions, and the truth may well lie between the two extremes.

The Hydrogen Molecule.—One of the most complete applications of the wave mechanical procedures in the study of molecular structure is to the interaction between two hydrogen atoms. If these two atoms are indicated by subscripts a and b , and the electrons associated with them are represented by (1) and (2), then the two atomic orbitals are $\psi_a(1)$ and $\psi_b(2)$, so that a permissible wave function of the hydrogen molecule, according to the homopolar method of Heitler and London, is $\psi_a(1)\psi_b(2)$. It is possible, however, to exchange the two electrons, so that the atomic orbitals are now $\psi_a(2)$ and $\psi_b(1)$, and the wave function of the molecule is $\psi_a(2)\psi_b(1)$. Since the electrons are identical, the energies must be the same, and as there are two possible orbital wave functions corresponding to this energy, more satisfactory solutions of the wave equation of the molecule can be obtained by linear combinations of these functions, viz.,

$$\Psi_S = \frac{1}{\sqrt{2 + 2T^2}} \{ \psi_a(1)\psi_b(2) + \psi_a(2)\psi_b(1) \} \quad (113)$$

and

$$\Psi_A = \frac{1}{\sqrt{2 - 2T^2}} \{ \psi_a(1)\psi_b(2) - \psi_a(2)\psi_b(1) \}, \quad (114)$$

where T , required for normalization, is a function only of the distance between the two hydrogen atoms. The functions Ψ_S and Ψ_A are called symmetric and antisymmetric, respectively, because a change of the electron coordinates leaves the eigenfunction unaffected in the former case but alters the sign in the latter. The wave function of an electron in the ground state of the hydrogen atom can be obtained, in terms of its distance from the nucleus, by solving the wave equation (101); hence, it is possible to derive expressions for the molecular wave functions given by (113) and (114). Insertion of these in the appropriate equation gives the corresponding eigenvalues of the energy as, approximately,

$$E_S - 2E_H = (Q + J)/(1 + T^2), \quad (115)$$

$$E_A - 2E_H = (Q - J)/(1 - T^2), \quad (116)$$

where E_S and E_A are the energies of the molecules with symmetric and antisymmetric orbital wave functions, respectively, and $2E_H$ is the energy of the two separate hydrogen atoms; (115) and (116) thus give the potential energies of the molecule, for various internuclear distances, with respect to the separated atoms. The negative of the potential energy is consequently a measure of the energy required to break up the molecule into its constituent atoms.

It is seen from (115) and (116) that the potential energy contains two terms, viz., Q and J , and both vary with the distance between the nuclei; Q is called the coulombic energy, due to classical electrostatic forces between electrons and protons, and J is the resonance or exchange energy (cf. p. 80). The values of Q can be calculated, but a precise estimate of J is difficult; the results, however, show that the potential energies given by (115) and (116) can be represented graphically by the two curves in Fig. 21. The hydrogen molecule with the antisymmetric orbital wave function is clearly unstable, since the potential energy is always greater than that of the separated atoms, as these are brought together there is a continuous increase in the repulsive force between them. When the orbital portion of the wave function is symmetric, however, there is

obviously a state of maximum stability, as indicated by the minimum of the potential energy curve. As the two hydrogen atoms are brought together, so as to form the "symmetric" molecule, they attract each other, thus leading to the formation of a stable state; if the distance of separation is made less than this equilibrium position requires, then repulsion sets in. It is evident, therefore, that the minimum of the curve will give the internuclear distance in a normal molecule of hydrogen and that the corresponding potential energy, with its sign reversed, should be equal to the heat of dissociation of the molecule into atoms. From the first approximate estimates of the resonance energy the distance between the hydrogen nuclei in a stable molecule was calculated to be 0.80 Å. and the heat of dissociation was estimated as 66.3 kcal. per mole, which may be compared with the experimental values of 0.74 Å. and 103.0 kcal., respectively, derived from the spectrum of molecular hydrogen (Chapter VIII).

The eigenfunctions (113) and (114) employed in the Heitler-London treatment are admittedly approximate, and various attempts have been made to obtain improved wave functions. Some of these involve the use of atomic orbitals, with the application of corrections for the effective nuclear charge and for the mutual polarization (distortion) of the electronic fields of the two hydrogen atoms. Other treatments employ molecular orbitals, and the striking, although highly involved, calculations of H. M. James and A. S. Coolidge (1933), based on this procedure, led to values of the internuclear distance and heat of dissociation of molecular hydrogen in almost exact agreement with experiment.

Wave Mechanics and Covalence.—In terms of wave mechanics the Pauli exclusion principle (p. 80) implies that in no atomic or molecular system can the complete electronic wave function of the system be symmetric. The complete wave function for an electron is the product of the orbital function, as developed above, which gives the probability of finding an electron at a given point, and the spin function; for the complete function to be antisymmetric, and therefore possible, one partial function must be symmetric and the other antisymmetric. It has been seen that a stable molecule of hydrogen can result only when the orbital function is symmetric, and so it follows that the electronic spin function must be antisymmetric. In other words, two hydrogen atoms can combine to form a molecule only when the two electrons have antiparallel, or oppositely directed, spins. This deduction was generalized by Heitler and London, who postulated that the fundamental basis of the electron duplet, proposed by Lewis to represent the covalent bond on empirical grounds, is the desire for electrons with unpaired spins to couple with each other. An

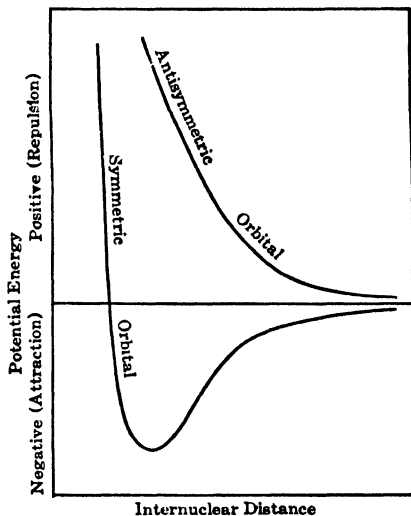


FIG. 21. Potential energy curves of molecular hydrogen (Heitler and London)

unpaired electron in one atom may be supposed to couple with a similar electron of opposite spin in another atom, thus forming a covalent bond, such as exists in hydrogen. The covalence of any element, not including dative covalences, should consequently be equal to the number of electrons with unpaired spins in the given atom. If the covalence is represented by V , then the resultant spin quantum number S of the atom must be $\frac{1}{2}V$, since each uncoupled electron contributes $\frac{1}{2}$ unit. The multiplicity of each line in the atomic spectrum is $2S + 1$ (p. 68), and so this multiplicity, represented by the superscript in the term symbol, should be equal to $V + 1$.

The nitrogen atom in its ground state has a 4S term (see Table 7, p. 86), and so it should be trivalent, in agreement with its chemical behavior. The normal state of carbon is 3P with two uncoupled electron spins; the covalence should, therefore, be two, although this value occurs in extremely few carbon compounds. It appears that the common quadrivalent compounds are derived from an excited 5S state in which the four electrons in the L -shell all have uncoupled spins, so that a valence of four is possible. It was believed at one time that relatively little energy was required to convert carbon from its normal 3P state to the 5S condition, which is apparently the first stage in the formation of quadrivalent derivatives; this view has been challenged and the situation appears to be obscure. It is nevertheless certain that the energy liberated in the formation of four bonds with a carbon atom (p. 590) is amply sufficient to provide the necessary excitation energy, even if it is as high as 7 e.v., as has been suggested (J. H. Van Vleck, 1934). In the 5S carbon atom the four electrons must have the configuration sp^3 if the Pauli principle is not to be disobeyed; there cannot be two s electrons, for their spins would be coupled, and there cannot be four in the p level, since the maximum in this subgroup is six and so not more than three such electrons can have uncoupled spins. In spite of this difference in the electrons, viz., one s and three p , the four covalences of carbon are identical. It appears that under certain conditions the quantization can be changed, as explained below, so that the s and p electrons have equivalent orbitals. This process is referred to as sp hybridization, and it has been suggested that it also occurs in the formation of the positively charged nitrogen atom which is the basis of the ammonium ion.

Directed Valence Bonds.—By solving the wave equation (101) for the hydrogen-like atom, it is possible to derive an expression for the "spherical" portion of the orbital wave function, which gives the angular distribution (or probability) of the electron. This can be evaluated for various quantum numbers l and m , and the results can be plotted in terms of three rectangular coordinates, x , y and z . In this way it is found that the s wave function ($l = 0$) is spherically symmetrical, having the same magnitude, at a given distance from the origin, in every direction. The p orbital ($l = 1$), however, takes the form of two spheres in contact, somewhat like a dumbbell or hourglass. There are three p orbital wave functions possible, corresponding to m values of 1, 0 and -1 , and these are identical in shape, but differ in orientation; each lies with its length along one of the three rectangular axes, so that they are at right angles to each other. The three p functions are consequently represented by the symbols p_x , p_y and p_z , respectively. The physical significance of these results is that the probability of finding an s electron is the same in all directions around the nucleus of a hydrogen-like atom, but the p electrons will tend to con-

centrate in certain preferred directions, these being at right angles to one another for the three types of p electrons in the atom.

In the **method of directed valence bonds**, developed by L. Pauling (1931) and J. C. Slater (1931), it is supposed that when two atoms are joined by a covalent bond, one electron being supplied by each atom, the direction of the bond will correspond to that in which the orbital wave functions of the two electrons overlap as much as possible. If this is the case, and the properties of the hydrogen-like s and p orbitals are assumed to be of general applicability, it follows that s electrons will have no directional effect, but p electrons will tend to form bonds making an angle of 90° with each other. In the elements of the oxygen and nitrogen groups bond formation takes place through the p electrons, since the s electrons have their spins paired; it is to be expected, therefore, that the bond angles in the resulting compounds will be about 90° . Apart from water and ammonia, where the mutual repulsion of the hydrogen atoms, due to their close proximity, increases the bond angles to 104° and 108° , respectively, the experimental values have been found to lie mainly between 92° and 97° .

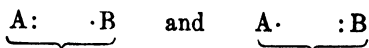
When sp hybridization occurs, as in carbon and positively charged nitrogen, the form of the orbitals is changed. For one s and three p electrons the wave mechanical calculations show that there are then four equivalent maxima, directed toward the corners of a regular tetrahedron. This result provides a theoretical basis for the familiar tetrahedral carbon atom and for the nitrogen atom in ammonium compounds. Another possibility is that hybridization may occur between one s and two p electrons, the third p electron being unaffected; the three equivalent orbitals, known as trigonal wave functions, are now found to have their maxima lying in one plane, the directions making angles of 120° with one another. If the two p orbitals are p_z and p_y , the three hybridized trigonal orbitals will lie in the xy plane, while the unaffected p_x orbital will be oriented in a direction normal to this plane. When two carbon atoms are joined by a double bond, it appears that one of the two linkages, sometimes called a $\sigma\sigma$ bond, is formed by the overlapping of one trigonal wave function from each atom, while the second ($\pi\pi$) linkage results from the pairing of the two pure p_z orbitals. In the $\sigma\sigma$ bond the orbitals of the two carbon atoms are colinear, but in the $\pi\pi$ bond they are parallel; the latter might thus be expected to be weaker than the former. This is in agreement with the fact that the energy required to break a $C=C$ bond is about 1.7 times that for a $C-C$ bond (cf. p. 590). Trigonal bond functions are probably involved in the benzene ring and in graphite, thus accounting for the characteristic carbon bond angle of 120° in these substances.

Molecular Orbitals and Valence.—In the interpretation of valence based on molecular orbitals, the concept of definite bonding pairs of electrons, and of directed valence bonds, is discarded; the particular stable structure of a molecule is then the arrangement of atomic nuclei with the minimum potential energy. It is assumed that there are definite electronic levels in a molecule corresponding to those in an atom, and if it is imagined that the electrons are added one by one to the system of nuclei, the orbital groups may be supposed to fill up gradually starting from the levels of minimum potential energy. If, as a result of lack of accommodation, an electron is forced to occupy in the molecule a principal quantum level higher than that in the atom whence it came, the electron is said to be **promoted**. In view of the large amount of en-

ergy required to bring about promotion such electrons tend to confer instability on the molecule and are said to be **antibonding**. When the electrons have a lower potential energy in the molecule than in the atom they are called **bonding electrons**, while others, particularly the two s electrons in the lowest, or K -, level, are **nonbonding**. In the nitrogen molecule, for example, there are four nonbonding electrons, arising from the $1s^2$ shells of the two atoms, two antibonding and eight bonding electrons. According to the molecular orbital theory all the electrons are involved in the formation of the nitrogen molecule. Since each molecular orbital, as mentioned above, is the resultant of the action of all the other electrons and the nuclei, it is apparent that all fourteen electrons are concerned in the formation of the molecule, but that no electrons are specifically involved in the bond. It may be assumed as an approximation, however, that the four nonbonding electrons play no part, and that the two antibonding cancel the effect of two of the bonding electrons, leaving six electrons for bonding purposes; this brings the molecular orbital theory into line with the duplet theory and the classical idea of the tervalence of nitrogen. It is important to emphasize, however, that this is not in strict harmony with the molecular orbital view; instead of regarding the molecule as made up of atoms, or ions, held together by valence bonds, it treats the molecule as a single unit built up of nuclei and electrons. It is not possible to associate different molecular orbitals with different atoms, for in general all the atoms participate in the construction of a given orbital. According to R. S. Mulliken (1931) the single bonding electron is the unit of valence, with the antibonding electron as a negative unit. It is an undoubted fact, however, that in compound formation electrons do tend to form pairs, but this is attributed solely to the necessity for the Pauli principle to be obeyed and the duplets are supposed to play no part in actual chemical bonding.

Odd-Electron Bonds.—It may appear, at first sight, that the molecular-orbital concept permits the possibility of two atoms being held together by a single electron, but this can only occur when there is approximate symmetry of the potential field with respect to the two atoms concerned. In these circumstances there is no tendency for the electron to favor one atom or the other, and consequently wave mechanical considerations require the inclusion of resonance energy which leads to a bonding effect. This point of view is virtually identical with one proposed by L. Pauling (1931) to determine the conditions under which two atoms might be expected to be held together by a single electron bond or, in general, by an odd-electron bond. The hydrogen molecule ion, H_2^+ , has a dissociation energy of 60.0 kcal., determined from spectroscopic measurements; this means that its one electron must have a considerable bonding effect. The explanation is that two equivalent states of the hydrogen molecule-ion, with the same energy, are possible; they are $\underline{H \cdot} \ H^+$ and $H^+ \cdot \underline{H}$, since the one electron is just as likely to be associated with one atom as the other. This possibility of two states of identical energy leads, as in analogous cases, to stabilization as a result of wave mechanical resonance. The one-electron bond, or **singlet linkage**, as it has been often called, is thus only possible when there are two conceivable electronic states of the same or similar energy. Many instances of supposed singlet linkages quoted in the literature do not satisfy this requirement, and so probably do not involve true odd-electron bonds.

The concept of the one-electron bond has been extended by Pauling to include three-electron linkages. If the atoms A and B forming a compound AB are similar, then the two configurations

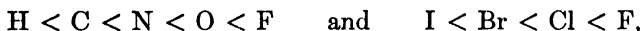


will have similar energies, and resonance will lead to the establishment of a stable three-electron bond. The structure of nitric oxide, for example, may be written $:N::O:$, since the nitrogen and oxygen atoms are so similar. The stable three-electron bond would explain the reluctance of nitric oxide to form N_2O_2 molecules, in spite of its odd number of electrons. To account for the heats of dissociation of three-electron linkages it is necessary to suppose that they have *half* the strength of a normal bond. According to the molecular orbital theory one of the electrons is probably antibonding in character and so the three electrons are virtually equivalent to one bonding electron only, and the relative weakness of the bond is explained.

The Oxygen Molecule.—The structure of oxygen has presented an important problem for the electronic theory of valence. There is no doubt, from spectroscopic measurements, that the two oxygen atoms are joined by a double bond, or its equivalent, but the simple structure $:\ddot{O}::\ddot{O}:$ is not in harmony with the paramagnetic properties of the gas. L. Pauling (1931) suggested the configuration $:\ddot{O}::\ddot{O}.$, with two three-electron bonds, which can be readily justified by the arguments given above. An odd-electron bond should confer paramagnetism, since the electrons must have a resultant spin momentum, and so the molecule, as written, should be paramagnetic. An alternative possibility, $:\ddot{O}::\ddot{O}:$, was proposed by R. F. Hunter and R. Samuel (1934), in which two odd electrons on the left-hand atom have uncoupled spins. Since it is improbable that these two electrons will prefer one oxygen atom to the other, it is reasonable to suppose that there will be continual interchange between the two atoms; this would lead to a structure essentially identical with the one involving three-electron bonds. It may be mentioned that the most satisfactory explanation of the paramagnetism of oxygen is reached by the molecular orbital method; it can be shown that on the gradual addition of sixteen electrons to two oxygen nuclei, the formation of a stable molecule requires the last two electrons to have spins which are not coupled to zero. The paramagnetic susceptibility of oxygen gas has the value to be expected for two uncoupled electron spins (see p. 618).²⁷

Polar Molecules.—The molecular orbital point of view is probably the more correct fundamentally, but it is not so convenient for chemical purposes as the explanation of valence in terms of definite bonds, generally involving electron duplets; although the interpretation may perhaps only be an approximation to the truth, it will be adopted as the basis for the discussion of valence problems. When a bond is formed between two identical atoms, viz., $A:A$, the electron pair may be regarded as shared equally between the two atoms. That is to say, the orbital wave function

may be regarded as consisting of homopolar terms only, or if ionic terms are included then the functions for the two possible ionic forms, viz., A^+A^- and A^-A^+ , are present to an equal extent. On the other hand, in bond formation between unlike atoms, e.g., $A:B$, it is a matter of experience that one atom is usually more electronegative than the other; that is, the electrons of the duplet are more likely to be found in the vicinity of one atom than the other. In the language of wave mechanics, the electronic part of the molecular wave function includes both homopolar and ionic terms, but the coefficients of the terms corresponding to the forms A^+B^- and A^-B^+ are not the same. If A is more electronegative than B , i.e., A has a greater tendency to attract electrons than B , then the contribution of A^-B^+ will be the larger. The important consequence of the different electron affinities of various elements is that most covalent bonds constitute electrical dipoles; the electrons may be regarded as drawn more to one atom than to the other, so that the electrical center of the electrons is not the same as that of the nuclei, the latter being virtually at the mass center of the molecule. In other words, a covalent bond between two different atoms will be associated with a **dipole moment**; its value is given by the product of the electronic charge and the relative displacement of the positive and negative electrical centers. The magnitude of the dipole moment, or of the **polarity** of the bond, to use a general term, depends primarily on the difference in electronegativity of the two elements forming it. This difference can be estimated directly by measurements of dipole moment (see Chapter VIII), or by comparing the bond energies as obtained from thermal data with those to be expected if the linkage were purely covalent. It appears from these considerations that the order of electron attraction is the following:



and similarly for other series and groups in the periodic table. In the formation of a bond the polarity increases with increasing separation of the atoms in the electronegativity series, the element to the left becoming positive with respect to the other.

There is little doubt that dipole moments of bonds are to be represented, approximately at least, as additive vector quantities acting in the direction of the chemical bond, and so the moment of the whole molecule is the vector sum of the constituent moments (p. 552). A perfectly symmetrical molecule will therefore be nonpolar, although it may contain polar linkages. A simple case of this kind is carbon tetrachloride, in which the moments of the four tetrahedrally directed $C-Cl$ bonds cancel each other. Unsymmetrical compounds, however, are almost invariably polar; the presence of oxygen, nitrogen or a halogen atom in a carbon compound makes this polarity very marked. Nonpolar substances are, as a rule, readily volatile and are soluble in other nonpolar compounds, whereas polar molecules, having an appreciable electrostatic force field, will attract each other, and the liquids will consequently have relatively

high boiling points. The physical properties of an apparent nonpolar molecule often depend on the distance between the constituent dipoles; if they are close together, as in carbon tetrachloride, the substance has a typically nonpolar character, but if they are far apart, as in *p*-dichlorobenzene, polar properties become evident although the molecule as a whole is theoretically nonpolar. It may be noted that in addition to polar and nonpolar linkages, and polar and nonpolar molecules, it is often the practice to speak of polar and nonpolar groups; an alkyl radical, for example, is generally regarded as nonpolar, although it has a small polarity, but $-\text{OH}$, $-\text{CN}$, $-\text{COOH}$, $-\text{NO}_2$, etc., are said to be polar groups.

It was seen on p. 100 that a dative covalence should be accompanied by an electrical dipole; such a bond generally produces a relatively large dipole moment, because the displacement between positive and negative electrical centers is of the order of an atomic radius, and the molecules are strongly polar, e.g., nitro-compounds and isonitriles. Nitro-compounds are consequently more polar than the isomeric nitrites, and the former have the higher boiling points; nitro-compounds are also less volatile than the analogous halogen compounds of the same or greater molecular weight. Polar compounds containing a dative bond often have high dielectric constants; this is connected with their large dipole moments (see Chapter VIII). It is important to note that in some instances the dipole moment due to a semipolar bond may be compensated by the opposing influence of the relative electronegativities of the elements; such is apparently the case with carbon monoxide. Its structure is probably to be represented by a combination of the two forms $\text{C}\equiv\text{O}^+$ and $\text{C}=\text{O}$; since oxygen is more electronegative than carbon the normal dipole moment of the $\text{C}=\text{O}$ linkage would make oxygen negative, and this counteracts the opposing influence of the dative bond. Carbon monoxide has consequently only a small dipole moment, and its physical properties, which resemble very closely those of nitrogen, are typical of a nonpolar gas.

Electrovalence and Homopolar Terms.—The extreme type of polarity arises when the bond between the atoms is an electrovalence; one of the ionic terms is now predominant in the orbital function of the molecule, the homopolar term and the other ionic contribution being relatively small. Such substances have, as is to be expected, high boiling points and are good conductors of electricity in the molten state or in solution. The work of K. Fajans (1924 *et seq.*) on the polarizability of ions (p. 540) suggests that with decreasing size and increasing charge of the positive ion, and with increasing size of the negative ion, the contribution of homopolar terms becomes more significant; this is accompanied by a decrease of boiling point and of electrical conductance in the fused state. Other nonpolar properties, such as solubility in organic liquids, appear at the same time. The rarity of quadrivalent cations is explained partly by the difficulty of removing four electrons from an atom, and partly by the

marked tendency for a highly charged positive ion to favor the formation of a covalent bond. Since the bare hydrogen ion, i.e., the proton, is very small, its presence in an ionizable compound, e.g., an acid, will result in the inclusion of considerable homopolar terms in the wave function. In the pure state such compounds, e.g., hydrogen chloride, either gaseous or liquid, are almost entirely covalent and are nonconductors. In certain solvents, however, ionization takes place, but this is probably accounted for by interaction with the solvent (cf. p. 975).

Coordination Compounds.—For metals of the alkali and alkaline-earth groups electrovalent compounds are by far the most common, since the ionization potentials are relatively small. In the third group, however, the homopolar influence becomes apparent and the linkages are to a great extent covalent in character; aluminum chloride, for example, has a low boiling point, viz., 183°C ., and the conductance of the liquid at its melting point is very small. It is true that ionization occurs in aqueous solution, but the ion is then probably a hydrated one, e.g., $\text{Al}(\text{H}_2\text{O})_6^{+++}$, the heat of the hydration process supplying the energy necessary for removing the three electrons to yield a trivalent ion. Metals belonging to the transition series, e.g., chromium, manganese, iron, cobalt, nickel, copper and zinc, appear to behave in a similar manner. The loss of one, two or three electrons does not give these elements the inert gas structure, and so there is little inclination for such ions to be produced, although they may exist to a small extent in solution. Elements of this kind are, however, characterized by the ease with which they form coordination compounds. It may be supposed that in these substances the coordinated groups are attached to the central atom by covalences, generally of the dative type if the group is a neutral molecule, e.g. NH_3 , $\text{NH}_2\text{CH}_2\text{CH}_2\text{NH}_2$, CH_3CN , etc. (N. V. Sidgwick, 1923). In the platinumous compound $\text{Pt}(\text{NH}_3)_4\text{Cl}_2$, for example, which dissociates into three ions in solution, the group $[\text{Pt}(\text{NH}_3)_4]^{++}$ is probably a bivalent cation; it may be regarded as formed by the donation to the platinum of a lone pair of electrons from each of the nitrogen atoms of the four ammonia molecules, with the subsequent loss of two electrons to form the bivalent positive ion. The total number of electrons associated with the platinum (atomic number 78) would thus be $78 + 8 - 2$, i.e., 84, which is presumably a stable group. If one of the ammonia molecules is replaced by a chlorine atom, which can only contribute one electron towards the sharing of a pair, there are $78 + 7$, i.e., 85 electrons associated with the central atom, and hence the loss of one electron, leading to the univalent cation $[\text{Pt}(\text{NH}_3)_3\text{Cl}]^+$, as found in practice, should give a stable electronic arrangement.* By the same argument it is easy to understand why the compound $\text{Pt}(\text{NH}_3)_2\text{Cl}_2$ is a nonelectrolyte. Proceeding in an analogous manner it is seen that to maintain the group of 84 electrons the radical $[\text{Pt}(\text{NH}_3)\text{Cl}_3]^-$ must be a univalent anion, and $[\text{PtCl}_4]^{--}$ should be

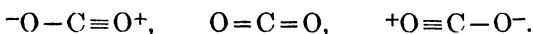
* The same conclusion is reached by assuming that a chloride ion donates a pair of electrons.

bivalent, as is found in K_2PtCl_4 . Crystal structure evidence (Chapter V) confirms the view that the coordinated groups are attached to the central atom by covalent linkages, as postulated above; there is, however, reason for believing that in some instances ionic terms may be important. A case of this kind is the ion FeF_6^{---} , in which magnetic measurements indicate that the union between the iron and fluorine is to a great extent ionic in character (p. 621); this is probably due to the strongly electronegative nature of the fluorine atom. In $Fe(CN)_6^{---}$, on the other hand, the bonding is almost entirely covalent, for the CN group is much less electronegative.²⁸

The Effective Atomic Number.—The number of electrons associated with the central atom in a stable coordination compound, e.g., 84 in the platinumous compounds, is called the **effective atomic number**, abbreviated to E.A.N. In general it is ascertained by adding to the number of electrons on the atom, i.e., its atomic number, twice the coordination number, since every neutral coordinated group donates two electrons, and subtracting the ordinary valence of the central atom. In the platinumous series considered above, for example, the atomic number is 78, the coordination number is 4, and the valence 2, so that the E.A.N. is $78 + 8 - 2$, i.e., 84. For the corresponding platonic compounds, the coordination number is 6 and the valence 4, so that the E.A.N. is 86; it is of interest that this figure is identical with the number of electrons in the inert gas radon. It is true for many other coordination compounds that the E.A.N. is the same as the atomic number of the corresponding inert gas, but there are other stable series for which this is not the case, although the discrepancy is rarely more than two. It is quite certain, however, in spite of the coincidence, that the electronic configurations of coordination compounds cannot be the same as those of the inert gases, for the latter have only eight electrons in their outermost shells and it is improbable that this will be the case in the former.

Resonance and Structure.—One of the most significant developments of the concept of wave mechanical resonance, due mainly to L. Pauling (1933 *et seq.*), is in connection with the phenomenon which has become known by the general name of **resonance**. If the electronic structure of a molecule can be depicted in two or more ways in which the positions of the atoms remain unchanged, and the energies of the various states do not differ appreciably from each other, then the actual state of the molecule has a greater stability, i.e., a greater negative potential energy, than any of the individual structures. The greater strength of the linkages, due to resonance, is accompanied by a shortening of the bond distances. It is possible in a few cases, by solving the wave equations, to calculate the magnitude of the resonance energy, but in most instances experimental methods are employed to detect resonance. In the first place, the increased potential energy will mean an increase in the heat of dissociation or heat of formation (cf. p. 573), and in the second place the interatomic distance will be less than if resonance did not occur.

The heat of formation of carbon dioxide from its constituent *atoms* is 336 kcal. per mole, whereas the value calculated on the assumption of the additivity of heats of linkage is about 300 kcal., i.e., twice that for a >C=O bond (p. 590). Further, the carbon-oxygen distance in carbon dioxide is 1.15 Å., compared with about 1.22 Å. to be expected for a double bond. It is probable, therefore, that resonance occurs between three possible structures, viz.,



The middle formula is the one based on classical ideas of valence, whereas the other two are equally possible structures; all three will have similar energies, and the effect of resonance is to give a hybrid structure which is a combination of, but more stable than, those depicted. It will be seen that the first and third forms should have high dipole moments, but they act in opposite directions and since the structures are identical it is evident they will contribute equally to the actual state of the molecule; the latter has therefore a dipole moment of zero. Observations of the bond distances in nitrate and carbonate ions show that resonance occurs among three equivalent structures in each case. The conven-

tional structure of the carbonate ion would be $\text{O}=\text{C}\begin{matrix} \text{O}^- \\ \diagup \\ \text{O}^- \end{matrix}$, but this cannot

be correct, as there is no difference between the three oxygen atoms. It is clear that three identical formulae are possible, in which each of the three oxygen atoms is doubly bound to carbon, and consequently the actual structure is a combination of them, all the carbon-oxygen distances, viz., 1.31 Å., being similar to those expected for a double bond, about 1.28 Å. The structure of the nitrate ion is based on similar considerations (see p. 392).

An especially interesting example of resonance is provided by the benzene molecule: five different and independent (canonical) structures of similar energy are possible, viz.,



I



II



III



IV

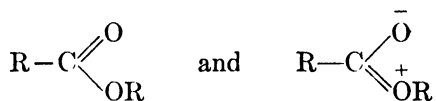


V

These structures represent five different, and independent, ways in which the six p_z orbitals of the carbon atoms can be joined in pairs, that is, to form the $\pi\pi$ bonds (cf. p. 105). The structures III, IV and V are less stable than the Kekulé structures I and II, because the former involve interactions of the p_z orbitals over unusually large distances. The heat of formation of benzene indicates a resonance energy of about 40 kcal. per mole (p. 590), and the distances between adjacent carbon atoms are all the same (1.40 Å.) and approach the normal double bond value (1.35 Å.). These facts clearly suggest the existence of resonance; the

actual structure of benzene, which cannot be represented in a pictorial manner, is a combination or hybrid of the five possible structures. Wave mechanical calculations imply that the contributions of forms I and II are far more important than those of the other three. Thermal measurements show that resonance occurs in aromatic compounds generally, the resulting stabilization energy being about 39 kcal. per benzene ring in each case. The exceptional stability of the aromatic nucleus and the absence of the chemical properties usually accompanying unsaturation are thus readily explained by the resonance structure of benzene and its derivatives.

In carboxylic acids, esters and amides there is the possibility of resonance among structures of the type



and the heats of formation indicate that this does actually occur, the resonance energy being about 25 kcal. per mole in each case. It has been suggested that the failure of these substances to exhibit the properties associated with the >C=O group in ketones and aldehydes is to be attributed to the stable resonance structure.²⁹

It should be pointed out that when resonance occurs the actual structure is not to be regarded as being represented by one or the other of the various configurations; nor is it to be considered as involving a rapid oscillation or interchange from one structure to another. The situation is that the actual electron distribution in the molecule cannot be represented in terms of a single structural formula, but it can be approximated by treating it as an appropriately weighted combination of several reasonable configurations.

The Hydrogen Bond.—For many years there has been growing evidence that in certain cases a hydrogen atom is able to act as a link between two electronegative atoms. The possibility of this phenomenon was first considered by T. S. Moore and T. F. Winmill (1912) and by P. Pfeiffer (1913), and its wide applicability was realized by W. M. Latimer and W. H. Rodebush (1920). In more recent times the concept of what was once called "coordinated hydrogen" or "hydrogen bridge," and now known as the **hydrogen bond**, has received general acceptance. An outstanding example of hydrogen bonding is to be found in the stable hydrogen fluoride ion, HF_2^- ; in view of the highly electronegative character of the fluorine atom, it is probable that the structure is essentially that of a proton holding two fluoride ions, viz., $\text{F}^--\text{H}^+\text{F}^-$, by electrostatic forces. The presence of a hydrogen bond accounts for the phenomenon of **molecular association**, that is, the formation of relatively complex molecular species by the union of two or more simple molecules, and also for

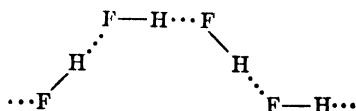
many unusual physical properties of water, liquid ammonia, alcohols, phenols, etc.

According to the Pauli principle a hydrogen atom can be associated with no more than two electrons, and hence in the formation of a hydrogen bond it is not possible for both atoms which are linked in this manner to be attached to the hydrogen by an ordinary covalence. Because of the experimental fact that the linked atoms must be strongly electronegative, it is probable that the hydrogen bond is largely electrostatic or ionic in nature. The strength of the bond increases, in general, with increasing electronegativity of the atoms which are bridged by the hydrogen; these atoms may be the same or different, but they are restricted almost exclusively to fluorine, oxygen and nitrogen (cf. p. 108), e.g., $\text{O}-\text{H}-\text{O}$, $\text{F}-\text{H}-\text{N}$, etc. In certain instances hydrogen bonds are formed between carbon and oxygen and between carbon and nitrogen; in such cases, however, the carbon atom is always attached to a strongly electronegative group. Although chlorine has approximately the same electronegativity as nitrogen, it is only infrequently involved in hydrogen bond formation; this is attributed to the larger size, and hence less intense field in the vicinity, of the former, i.e., chlorine, atom. However, chlorine and even sulfur are sometimes involved in a hydrogen bond, but the other atom must be more strongly electronegative, e.g., oxygen or nitrogen.

The general question of bond energies, i.e., the mean energy required to break a bond of a particular type, will be treated in Chapter VIII but it is of interest to mention here that hydrogen bond energies are about 2 to 8 kcal. per mole. depending on the electronegativity of the two elements concerned. This value is very small compared with those for ordinary valence bonds, e.g., 87 kcal. for $\text{C}-\text{H}$, 84 kcal. for $\text{N}-\text{H}$, etc. (see p. 59C), and so accounts for the relative ease with which hydrogen bonds are broken especially as the temperature is raised.

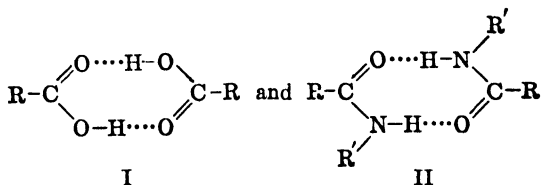
A comparison of the hydrogen compounds of fluorine, oxygen and nitrogen with those of their homologues in the periodic table reveals discrepancies in physical properties, although methane is normal in this respect. The melting and boiling points and the heats of vaporization of hydrogen fluoride, water and ammonia are much greater than might have been anticipated. The familiar fact that water is a liquid, while hydrogen sulfide is a gas at ordinary temperatures is an illustration of this type of abnormality. These results may be explained by hydrogen bond formation, leading to molecular association in both solid and liquid states of hydrogen fluoride, water and ammonia. The molecules are thus more complex than would be expected from the formulae HF , H_2O and NH_3 , respectively. Owing to the decreased electronegativity and larger size of the chlorine, sulfur and phosphorus atoms, the extent of hydrogen bond formation and of association in HCl , H_2S and PH_3 is negligible and these substances actually occur as single molecules. In hydrogen fluoride the strength of the hydrogen bond is relatively so large that association persists even in the vapor state; there is evidence

from electron diffraction measurements (cf. p. 591) that chains, e.g.,



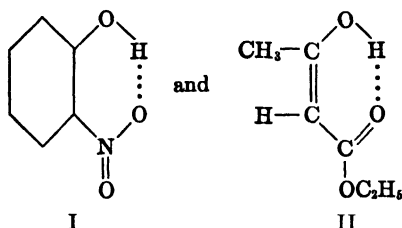
are formed, the $\text{H}-\text{F} \cdots \text{H}$ bond angle being about 140° . There is undoubtedly considerable association in water, for although the $\text{O}-\text{H} \cdots \text{O}$ bond is weaker than the corresponding bond involving fluorine, the fact that the oxygen atom in water has two lone pairs of electrons makes it possible for each such atom to form two hydrogen bonds, as well as two covalent bonds. In ice each oxygen atom is surrounded by four others (see p. 394), presumably with a hydrogen atom as a link between each pair. Thus every oxygen atom takes part in the formation of two hydrogen bonds, and the whole crystal of ice is virtually one molecule. A somewhat similar situation exists in liquid water, although the amount of hydrogen bond formation decreases with increasing temperature, and water vapor consists of single molecules. In liquid ammonia the extent of association is not large, for two reasons: first, the nitrogen atom is less electronegative than oxygen and fluorine, and second, it has but one lone pair of electrons, so that each nitrogen atom can be concerned in one hydrogen bond only. However, just as with hydrogen fluoride and water, the linking can extend from one molecule to another through the liquid.

Because of the formation of $\text{O}-\text{H} \cdots \text{O}$ bonds, all compounds of the ROH type, e.g., water, alcohols and phenols, are associated. Various aspects of this subject will be discussed in later chapters, but it may be emphasized that the association is of a general type, involving the linking together of a relatively large, but indefinite, number of single molecules by hydrogen bonds. It is evident that explicit polymers, such as $(\text{H}_2\text{O})_3$, $(\text{H}_2\text{O})_2$, etc., which were at one time postulated, do not exist, except perhaps instantaneously. With carboxylic acids and N-substituted amides, however, definite double molecules are formed; in the acids at least, these are known to be present in the solid and liquid, and even in the vapor if the temperature is not too high (p. 319). The structures of the double molecules, which depend on hydrogen bond formation, viz., $\text{O}-\text{H} \cdots \text{O}$ in the acids (I) and $\text{N}-\text{H} \cdots \text{O}$ in the amides (II), may be pictured as follows:



In the cases discussed so far the hydrogen bond has been between two similar atoms derived from different molecules of the same substance. Other types,

involving a single molecule only, or molecules of two different substances, are known. Intramolecular hydrogen bond formation, for example, accounts for the exceptional physical properties, such as melting point and solubility, of many *ortho*-substituted phenols, e.g., nitrophenol (I), and also of the enolic forms of β -keto-esters, e.g., ethyl acetoacetate (II); the structures are shown below.



The loose combination, or intermolecular association, sometimes referred to as the formation of "molecular compounds," between two substances in the liquid state can frequently be explained by the presence of a hydrogen bond. This is the case, for example, when a hydroxylic compound is mixed with another similar compound, or with an ether, ester or ketone; thus, a mixture of methyl alcohol and acetone probably contains complexes such as $\text{CH}_3\text{O}-\text{H}\cdots\text{OC}(\text{CH}_3)_2$. Other instances, in which the hydrogen links two dissimilar atoms, are also known.³⁰

REFERENCES

- Born, M. "Atomic Physics."
 Herzberg, G. "Atomic Spectra and Atomic Structure." (Translated by J. W. T. Spinks.)
 Hoag, J. B. "Electron and Nuclear Physics."
 Millikan, R. A. "Electrons, etc."
 Pauling, L. "The Nature of the Chemical Bond."
 Pauling, L. and Goudsmit, S. "The Structure of Line Spectra."
 Rice, O. K. "Electronic Structure and Chemical Binding."
 Richtmeyer, F. K. "Introduction to Modern Physics." (Revised by E. H. Kennard.)
 Ruark, A. E. and Urey, H. C. "Atoms, Molecules and Quanta."
 Stranathan, J. D. "The Particles of Modern Physics."
 Taylor, H. S. and Glasstone, S. "Treatise on Physical Chemistry," 3rd. ed., Vol. I. Chap. II (S. Dushman).
 Thomson, J. J. and Thomson, G. P. "The Conduction of Electricity Through Gases."
 White, H. E. "Introduction to Atomic Spectra."
1. Helmholtz, *J. Chem. Soc.*, **39**, 297 (1881) Faraday Lecture.
 2. For early history of studies of electrical phenomena in gases, see Townsend, "Electricity in Gases."
 3. Thomson (J. J.), *Phil. Mag.*, **5**, 346 (1903); Wilson (H. A.), *ibid.*, p. 429.
 4. Millikan, *Phys. Rev.*, **2**, 143 (1913); *Phil. Mag.*, **34**, 1 (1917); Hopper and Laby, *Proc. Roy. Soc.*, **178A**, 243 (1941).
 5. Compton and Doan, *Phys. Rev.*, **27**, 104 (1926); DuMond, *et al.*, *ibid.*, **50**, 524 (1936); **54**, 1005 (1938); **57**, 198 (1940); **58**, 457 (1940); Bearden, *J. App. Phys.*, **12**, 395 (1941).
 6. Davisson and Germer, *Phys. Rev.*, **30**, 705 (1927); Germer, *J. Chem. Ed.*, **5**, 1041, 1255 (1928); Thomson (G. P.), *Proc. Roy. Soc.*, **117A**, 600 (1928); "The Wave Mechanics of Free Electrons"; Clark and Wolthuis, *J. Chem. Ed.*, **15**, 641 (1938).

7. Anderson, *Phys. Rev.*, **43**, 491 1034 (1933); **45**, 352 (1934); Blackett and Occhialini, *Proc. Roy. Soc.*, **139A**, 699 (1933); Chadwick, *et al.*, *ibid.*, **144A**, 235 (1934); Spees and Zahn, *Phys. Rev.*, **58**, 861 (1940).
8. Rutherford, *Phil. Mag.*, **21**, 669 (1911).
9. van den Broek, *Phys. Z.*, **14**, 32 (1913).
10. Moseley, *Phil. Mag.*, **26**, 1024 (1913); **27**, 703 (1914); Bohr and Coster, *Z. Phys.*, **12**, 342 (1923); Coster and Hevesy, *Chem. and Ind.*, **42**, 258 (1923); Hevesy, "Chemical Analysis by X-Rays"; Compton and Allison, "X-Rays in Theory and Experiment."
11. Chadwick, *Phil. Mag.*, **40**, 734 (1920).
12. Chadwick, *Proc. Roy. Soc.*, **136A**, 692 (1932); **142A**, 1 (1933).
13. Neddermayer and Anderson, *Phys. Rev.*, **51**, 884 (1937); Bethe, *ibid.*, **57**, 260, 390 (1940); Marshak, *ibid.*, **57**, 1101 (1940).
14. Planck, *Verh. Deut. Phys. Ges.* **2**, 237 (1900); *Ann. Phys.*, **4**, 553 (1901); Einstein, *ibid.*, **17**, 132 (1905); **20**, 199 (1906); Dunnington, *Rev. Mod. Phys.*, **11**, 65 (1939).
15. Bohr, *Phil. Mag.*, **26**, 1, 476, 857 (1913).
16. Franck and Hertz, *Verh. Deut. Phys. Ges.*, **16**, 10 (1914); Davis and Goucher, *Phys. Rev.*, **10**, 101 (1917); Olmstead and Compton, *ibid.*, **22**, 559 (1923); Hertz, *Z. Phys.*, **18**, 307 (1923).
17. Sommerfeld, *Ann. Phys.*, **51**, 1 (1916).
18. Uhlenbeck and Goudsmit, *Nature*, **117**, 264 (1926).
19. Stern and Gerlach, *Z. Phys.*, **7**, 249 (1921); *Ann. Phys.*, **74**, 673 (1924).
20. Russell and Saunders, *Astrophys. J.*, **61**, 38 (1925).
21. Kossel, *Z. Phys.*, **1**, 124 (1920).
22. Heisenberg, *ibid.*, **38**, 411 (1926); **41**, 239 (1927); Darrow *Rev. Mod. Phys.*, **6**, 23 (1934); Pauling and Wilson, "Introduction to Quantum Mechanics"; Dushman, "The Elements of Quantum Mechanics."
23. Pauli, *Z. Phys.*, **31**, 765 (1925).
24. Thomson, *Phil. Mag.*, **7**, 237 (1904); Kossel, *Ann. Phys.*, **49**, 229 (1916); Lewis, *J. Am. Chem. Soc.*, **38**, 762 (1916); Langmuir, *ibid.*, **41**, 868, 1543 (1919); **42**, 274 (1920); Bury, *ibid.*, **43**, 1602 (1921); Bohr, "Theory of Spectra and Atomic Constitution" (1921); Main Smith, *Chem. and Ind.*, **42**, 1073 (1923); **43**, 323 (1924); Stoner, *ibid.*, **48**, 719 (1924).
25. Giaque and Johnston, *J. Am. Chem. Soc.*, **50**, 3221 (1928); Bonhoeffer and Har-teck, *Z. physikal. Chem.*, **B4**, 113; **B5**, 292 (1929); Eucken and Heller, *ibid.*, **B4**, 142 (1929); Farkas, "Orthohydrogen, Parahydrogen and Heavy Hydrogen"; Glasstone, "Theoretical Chemistry," Chap. IV.
26. Lewis, "Valence and the Structure of Atoms and Molecules"; Sidgwick, "The Elec-tonic Theory of Valency"; "The Covalent Link in Chemistry."
27. Heitler and London, *Phys. Z.*, **44**, 455 (1927); **46**, 455 (1928); Lennard-Jones, *Trans. Farad. Soc.*, **25**, 668 (1929); Pauling, *Chem. Rev.*, **5**, 185 (1928); *J. Am. Chem. Soc.*, **53**, 1367, 3225 (1931); **54**, 988, 3570 (1932); Slater, *Phys. Rev.*, **37**, 481; **38**, 1109 (1931); Mulliken, *Rev. Mod. Phys.*, **4**, 1 (1932); Van Vleck and Sherman, *ibid.*, **7**, 167 (1935); Glasstone, "Theoretical Chemistry," Chap. III.
28. Sidgwick, ref. 26; *J. Chem. Soc.*, **123**, 725 (1923); Sidgwick and Powell, *Proc. Roy Soc.*, **176A**, 153 (1940); Blanchard, *Chem. Rev.*, **26**, 409 (1940).
29. Pauling, *et al.*, *J. Chem. Phys.*, **1**, 362, 606, 679 (1933); Sidgwick, *J. Chem. Soc.*, 533 (1936); 694 (1937); Mills, *ibid.*, 340 (1944); see also, Glasstone, ref. 27.
30. Latimer and Rodebush, *J. Am. Chem. Soc.*, **42**, 1419 (1920); Rodebush, *Chem. Rev.*, **19**, 59 (1936); *J. Phys. Chem.*, **43**, 219 (1939); Huggins, *J. Org. Chem.*, **1**, 407 (1936); Sidgwick, *Ann. Rep. Chem. Soc.*, **30**, 112 (1933); **31**, 40 (1934); Lassettre, *Chem. Rev.*, **20**, 259 (1937); Pauling, "The Nature of the Chemical Bond," Chap. IX; General Discussion, *Trans. Farad. Soc.*, **36**, 851-928 (1940).

CHAPTER II

RADIOACTIVITY AND ISOTOPES

RADIOACTIVITY

Historical Introduction.—In attempting to discover a possible connection between X-rays and the luminescence observed in a discharge tube, H. Becquerel (1895) found that after exposure to cathode rays potassium uranyl sulfate possessed the property of affecting a photographic plate wrapped in black paper, indicating that the uranium salt was emitting a penetrating type of radiation. In the same year he made the surprising discovery that uranium compounds alone, without any previous treatment, are capable of fogging a photographic plate, and so emit rays spontaneously. In addition to their photographic action, the radiations were found, like X-rays, to be capable of ionizing the air, so that the activity of a uranium compound could be measured by the rate at which a known quantity caused the discharge of an electroscope. The emission of rays capable of producing these effects is a fundamental property of the uranium atom, as the rays are observed with various uranium salts in different valence states as well as with the element itself; further, the activity is found to be independent of the temperature or previous history of the material. The spontaneous emission of radiation of this type is now known as **radioactivity**, and a study of the subject has thrown much light on the structure of matter.

When examining the ionizing activity of the mineral *pitchblende*, one of the chief ores of uranium and consisting mainly of U_3O_8 , Mme. M. Curie and her husband, P. Curie (1898), noted that it had a greater activity than was expected from the uranium it contained. This result indicated the presence in the ore of compounds of an element, or elements, even more radioactive than uranium, and by using ordinary chemical methods of separation, two such substances were isolated. One of the elements was precipitated as its sulfide with bismuth sulfide; it was called *polonium*, in honor of Poland, the native country of Mme. Curie. The other element separated together with barium as their sulfates, which were subsequently converted into the bromides and separated by fractional crystallization. The element, obtained as the impure bromide by M. Curie, P. Curie and G. Bémont (1898), was given the name *radium*, because of its exceptional activity. About the same time, Mme. Curie and G. C. Schmidt (1898), independently, discovered that thorium compounds were also radioactive, and shortly afterwards, A. Debierne (1899) and F. Giesel (1901) found the new radioelement *actinium* in uranium minerals. Following these early discoveries, systematic investigation has

brought to light the existence in nature of more than forty elements which, either in the pure state or as compounds, emit characteristic radiations quite spontaneously and without any preliminary treatment. To these must be added the large number of radioclements not normally found in the earth, but capable of being generated by special methods to be considered later.

In the course of a study of the penetrating power of the radiations, E. Rutherford (1899) concluded that they could be divided into two types, which he referred to as α - and β -rays, respectively. The former were more easily absorbed than the latter but, on the other hand, they produced a much stronger ionization. Shortly afterwards P. Curie (1900) found that part of the radiation was not deflected in a magnetic field, and this was shown by P. Villard (1900) to have exceptional penetrating power, these radiations were called the γ -rays. As a rough approximation, they are about 10 to 100 times as penetrating as β -rays, which have a penetrating power about 100 times as great as α -rays. The three types of radiation may be most easily distinguished by the action of a magnetic field: the γ -rays are not affected at all, the β -rays are considerably deflected in such a direction as to indicate that they carry a negative charge, whereas the α -rays are deflected to a smaller extent in the opposite direction showing that they are positively charged.

The three kinds of rays also ionize the air to a different extent; in this respect the α -rays are the most effective being of the order of 100 times as powerful as the β -rays, and the latter approximately 100 times as active as the γ -rays. The ionizing power of the radiations has been utilized in two directions in the study of radioactivity and allied phenomena. If the active material is allowed to ionize the gas between two metal plates connected to a battery, a current can flow which increases to a maximum as the applied voltage is increased. This "saturation current," as the maximum value is called, is a measure of the number of ions present in the gas, for at high field strength all the ions formed travel to the charged plates with such high velocity that the extent of recombination of ions and electrons, leading to the formation of neutral particles, is negligible. The other application involves the use of the Wilson cloud-track apparatus described on p. 20; the path of radiations capable of producing ionization is made visible by the minute drops of water which condense on the ions.

After this general introduction to the subject of radioactivity, the properties of the three kinds of rays will now be considered in greater detail and their nature elucidated.¹

Alpha-Rays.—The deflection of α -rays in a magnetic field shows them to consist of positively charged particles, and it should be possible to determine the ratio of their charge to mass by studying the deviation of their path in superimposed electric and magnetic fields, as for electrons (p. 3). The earliest observations, by E. Rutherford (1903), gave an approximate value of 6×10^3 e.m. units per g. for e/m , but from re-

cent work it appears that a more accurate result is 4.813×10^8 e.m. units per g., and this is independent of the source of the α -particles. According to Faraday's laws (cf. p. 885) a g. ion of hydrogen carries 9649 e.m. units, so that the ratio of charge to mass is almost exactly twice that for an α -particle. If the latter carries a single charge, that is, the same as the hydrogen ion, then its mass must be twice as great, but if it carries two charges it has four times the mass of the hydrogen ion. In order to distinguish between these possibilities it is necessary to determine the magnitude of the charge carried by an α -particle. The methods described below indicate that this is approximately 9.3×10^{-10} e.s. unit, which is almost twice the charge on an electron. It follows, therefore, that the α -particle carries two unit charges, and hence it must have a mass approximately four times that of the hydrogen atom. In other words, the results suggest that the α -particle is a helium atom carrying two unit positive charges. As already seen (p. 30) this leads to the conclusion that the α -particle is a helium nucleus, and probably consists of two neutrons and two protons very firmly bound together.

Determination of Charge of Alpha-Particle.—The method employed by E. Rutherford and H. Geiger (1908) was to allow the α -rays from a given radioactive source to pass through an aperture of known area, at a known distance from the source, on to a plate attached to an electrometer. A magnetic field was used to deflect β -rays, and also electrons that might be produced by collision of the α -rays with matter, so that only the α -particles being studied fell on the plate. The charge acquired by the electrometer in a given time was measured, and from this it was possible to calculate the total charge carried by all the α -particles emitted from the radioactive source in unit time, assuming they were given off symmetrically in all directions. It is now necessary to know the number of α -particles produced by the given source in unit time; two chief methods have been devised to obtain this information. Rutherford and Geiger took advantage of the ionization produced by the α -particles. A gas at low pressure was placed in a vessel to the walls of which was applied a voltage somewhat below the sparking potential. No appreciable current passed until the entry of an α -particle resulted in the formation of ions, when there was a momentary deflection of the needle of a quadrant electrometer placed in the circuit. By means of a high resistance connected to earth the charge on the electrometer was able to leak away between the entry of successive α -particles. The radioactive material was placed at such a distance from the measuring apparatus that only three to five α -particles per minute entered it through a thin sheet of mica. By counting the number of deflections of the electrometer in a given time the rate of entry of α -particles could be determined, and hence the rate of emission in all directions from the given source could be evaluated. The method of counting α -particles was improved by E. Rutherford and H. Geiger (1912), using a string galvanometer and recording the deflections on a moving photographic film. A further modification was made by H. Geiger (1913) which could be employed for counting both α - and β -particles, and modern forms of the apparatus are based on that designed by H. Geiger and W. Muller (1928). The Geiger-Muller counter consists of a sealed cylinder, about 8 cm. long and 1.5 cm. diameter, with a thin wire of steel, aluminum or

tungsten stretched along its axis; this wire is made about 1000 volts positive with respect to the walls of the cylinder. The entry of any charged particle into the cylinder produces ionization of the air present at low pressure in it, and a momentary discharge takes place between the walls and the central wire. The effect is transmitted by a relay to a counter, and every particle entering the cylinder is recorded automatically in this manner.

The second method of counting α -particles utilizes the discovery, made independently by W. Crookes (1903) and by J. Elster and H. Geitel (1903), that the impact of α -particles on a screen covered with a special form of zinc sulfide produced a luminosity which proved, on magnification, to consist of a large number of small flashes or scintillations. This phenomenon was made the basis of the instrument called a **spinthariscopes** by Crookes; it consisted of a small tube, with a lens at one end and a zinc sulfide screen at the other, a trace of radioactive material being placed on a point at a short distance from the screen. Observation through the lens showed the appearance of numerous minute flashes of light at various points on the screen. The effect is apparently due to the bombardment of the zinc sulfide by the α -particles, each particle producing a scintillation on impact. By placing the radioactive material at a distance from the screen and limiting the exposed area, E. Regener (1908) found it possible, with the aid of a low-power microscope, to count the number of scintillations produced by a known amount of radioactive substance in a given time; from this the total rate of emission of α -particles could be calculated, assuming every α -particle produces a flash on the zinc sulfide. The method of direct counting has some advantages, especially as the effect of extraneous electrons and of β -particles is so minute as to be quite negligible, and γ -rays produce a general glow which does not consist of individual flashes. The results obtained were in good agreement with those given by the ionization method.²

Alpha-Particles and Helium.—Some confirmation of the view that the α -particle is a helium nucleus was found in the early observation of W. Ramsay and F. Soddy (1903) that helium is continuously being produced from radium, and the fact that radioactive minerals invariably contain the gas points in the same direction. More definite proof, however, of the identity of α -particles with charged helium atoms was obtained by E. Rutherford and T. Royds (1909). A quantity of the purified radioactive gas radon (see p. 127) was placed in a thin-walled tube A (Fig. 22), surrounded by an outer tube T attached to a spectrum tube S, through which an electric discharge could be passed. The tube T was evacuated at the commencement, and after several days the gas collected in it was

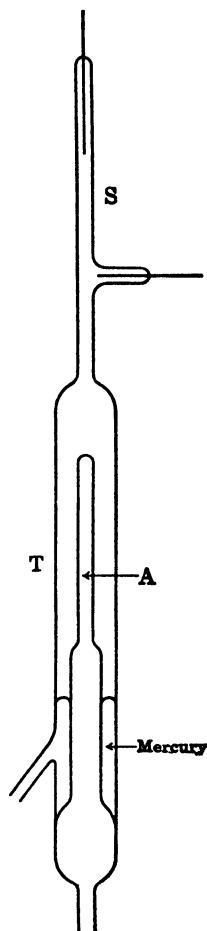


FIG. 22. Detection of helium in radioactive decay (Rutherford and Royds)

compressed, by raising the level of the mercury shown, into the tube *S*; on passing a discharge, the spectrum of helium was obtained. Independent experiments showed that ordinary helium gas could not have penetrated the walls of the tube, but α -particles are able to do so because of their high velocity. In this way the identity of these particles was established.

The Avogadro Number.—An interesting consequence of the experiments on α -particles is that the results permit of an independent determination of the Avogadro number (p. 1). The volume of helium gas liberated from radium in the course of disintegration, and presumably due to α -particles that have lost their charges, is 0.043 ml. per g. of radium per year, after making allowance for the α -particles produced by the disintegration products. The number of α -particles emitted in the same time by the same amount of radium is 11.6×10^{17} . Since each α -particle should presumably yield one helium atom or molecule, as this element has only one atom in the molecule, it follows that 0.043 ml. should contain 11.6×10^{17} molecules. A volume of 22.4 liters of helium, which is the volume occupied by 1 mole at S.T.P., would, therefore, contain

$$\frac{11.6 \times 10^{17} \times 22,400}{0.043} = 6.04 \times 10^{23} \text{ molecules.}$$

This value of the Avogadro number is very close to those obtained by other methods (see pp. 129, 263). It may be mentioned that this agreement, together with the fact that α -particles can be counted individually by their ionization effect or by the scintillations they produce, is to be regarded as strong evidence for the discrete, that is, atomic and molecular, structure of matter.

Velocity and Range of Alpha-Particles.—From the deflection of α -rays in electric and magnetic fields it is possible to determine the velocity of the particles constituting the rays, by a method similar to that used for electrons (p. 3). The results show that the initial velocity of an α -particle depends on the radioelement from which it originates, but most of, although not always all, the particles emitted by a given element have the same velocity; this is generally of the order of 1.4 to 2.0×10^9 cm. per sec., that is, about one-fifteenth to one-twentieth of the velocity of light. The initial velocity of an α -particle is related to its range in air, that is, the distance from its source which the particle travels in air before it ceases to have any appreciable power of producing ionization or scintillation. As a general rule these characteristic properties can no longer be observed when the velocity has fallen to about 40 per cent of its initial value in air. The initial velocities and ranges in air, at atmospheric pressure, at 15°C. for a number of radioelements are quoted in Table 16. It was shown by H. Geiger (1910) that the initial velocity v_0 cm. per sec. of an α -particle is related to its range R cm. in air by the equation

$$v_0^3 = aR, \quad (1)$$

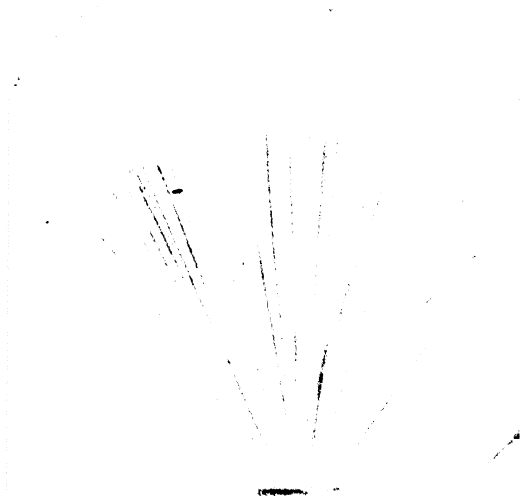
where the constant a has a value of about 1.0×10^{27} .

TABLE 16. INITIAL VELOCITIES AND RANGES OF α -PARTICLES

Radioelement	Velocity cm./sec.	Range cm.
Thorium	1.38×10^9	2.59
Uranium-I	1.39	2.63
Ionium	1.48	3.19
Radon	1.61	4.12
Thorium-A	1.80	5.68
Thorium-C'	2.05	8.62

The range of the α -particles from a given source varies with the nature of the medium through which they pass: it is said, therefore, that the **stopping power** for α -particles depends on the material. The stopping power increases with the atomic weights of the elements present, and as a rough approximation it is proportional to the square root of the atomic weight, when reduced to constant density.

Passage of Alpha-Particles Through Matter.—Since the α -particle has appreciable mass and moves with high velocity, it has considerable kinetic energy; it is thus very effective in removing electrons from molecules close to its path and so produces considerable ionization. For this reason α -particles give very distinct cloud tracks in the Wilson apparatus, and they can be very conveniently studied by this means. The photographs show that the majority of the particles travel through a gas in a

FIG. 23. Cloud-track photograph of α -particles from polonium

straight line, and the limited range is evident (Fig. 23). Occasionally a sudden deviation is observed at the end of the path which is probably due to a direct collision with a molecule of the gas; as a result of such collisions new particles are sometimes produced and their tracks can be observed.

In their passage through a thin sheet of metal, α -particles are scattered to a slight extent only; the majority of the particles travel in a straight line, or deviate little from a straight line, but a small proportion are violently deflected. The importance of this fact and some relevant data have been already discussed on p. 23, where it was shown that the observations led to the development of the nucleus theory of the atom.

Beta-Rays.—The deflection of β -rays in a magnetic field shows that they consist of rapidly moving particles carrying a charge of negative electricity, and measurements in magnetic and electric fields show that the ratio of charge to mass, for low velocities, is 1.77×10^7 e.m. units per g. This figure is almost identical with that obtained for electrons (p. 6), and so it is evident that a β -particle is actually a fast-moving electron. As with electrons, the effective mass of a β -particle varies with its speed (p. 7), and hence so also does its charge to mass ratio. The observed velocities of the β -rays depend, as do those of α -rays, on the radioactive element from which they originate, but the former differ from the latter in an important respect. Whereas almost all the α -particles from a given substance have the same velocity, the β -particles have velocities varying over an appreciable range. By means of a magnetic field it is possible to spread out β -rays from any definite source into a *continuous* β -ray spectrum, that is to say, there is a continuous variation of velocity. It is not easy to explain this effect, because the law of conservation of energy requires all the electrons from a given source to have the same kinetic energy, and hence a constant velocity. The suggestion has been made by W. Pauli, and developed by E. Fermi (1934), that the emission of a β -ray electron from an atom is accompanied by a neutral particle, that is one without charge, having a very small mass when at rest, probably less than that of an electron. This new particle has been called a **neutrino**, and it is supposed to be able to share with the electron the energy accompanying the radioactive change; a slow-moving β -particle would thus be associated with a fast neutrino. In view of its exceptional properties, viz., small mass and no charge, the detection of the neutrino, even if it does exist, will be very difficult. It possesses momentum, however, and so should produce a recoil effect; it is in this connection that some evidence for the neutrino has been obtained.

Velocity and Penetrating Power.—The velocities of β -particles approach that of light, an average value being 2×10^{10} cm. per sec., although, as already indicated, many from the same source may have higher or lower speeds. The highest observed velocity, about 0.99 times that of light, occurs with the most rapidly moving particles from radium-C'. Because of their high speed β -particles have considerable penetrating power, but on account of their small mass they are easily deflected from their course; they do not move in straight lines, therefore, but have very tortuous paths through matter. This fact is clear from the cloud-track photographs obtained when β -particles are allowed to enter a Wilson cloud-chamber (Fig. 24). For this reason it is impossible to measure the range of a β -particle. The small mass also accounts for the ionizing

power being less than for α -rays, although some β -particles which travel with high velocities possess energies and ionizing powers approaching those of α -particles.



FIG. 24 Cloud-track photograph of β -particles from radium-*E*. (The curvature of the tracks is produced by a magnetic field)

Gamma-Rays.- The failure of electric and magnetic fields to deflect γ -rays, and their exceptional penetrating properties, suggest that they consist of electromagnetic radiations analogous to X-rays. As with the latter (see p. 370), the wave lengths of γ -rays may be determined by using a crystal as a diffraction grating (E. Rutherford and E. N. da C. Andrade, 1914); the values obtained are even lower than for X-rays, wave lengths between 10^{-8} and 10^{-11} cm. having been recorded. The γ -rays are generally accompanied by secondary electrons, which are produced by their passage through matter, and from the kinetic and binding energies of these electrons it is possible to calculate, by means of the familiar quantum theory equation, $E = h\nu$, the wave lengths of the γ -rays. Even from a given radioactive element the values are not all the same, and definite γ -ray spectra have been obtained by using a crystal as a diffraction grating and employing a very small glancing angle to compensate for the low wave length (cf. p. 13).

The emission of γ -rays probably accompanies most radioactive changes, but the rays are not always observed. It was believed at one time that the emission of γ -rays occurred simultaneously with that of the α - or β -particle, but the work of L. Meitner (1925) has shown it to take place subsequently. After the α - or β -particle is discharged from the

atomic nucleus, there is often a rearrangement in the latter resulting in a liberation of energy, and it is apparently this which appears in the form of γ -rays.³

Radioactive Decay.—In addition to the discovery of the three types of radiation just described, the early workers in radioactivity brought other facts to light that have helped in the understanding of the complex phenomena. If a solution of a uranium compound, together with a ferric salt, is treated with ammonium hydroxide or ammonium carbonate, the precipitate is found to contain almost the whole of the original activity, whereas the liquid, which still contains the uranium in solution, is almost inactive (H. Becquerel, 1901). It appears, therefore, that most of the observed activity of the compounds of uranium is not due to that element, but to another separable from it to which the name uranium-X was given. On standing for some weeks, however, the uranium solution gradually regains its initial activity, but the precipitate loses its activity at the same rate (Fig. 25), so that the total activity remains constant. The

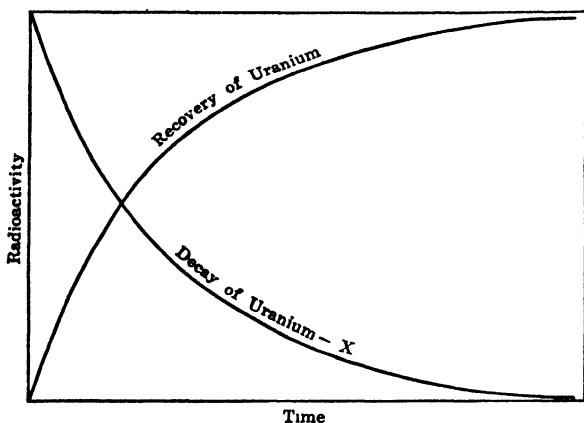


FIG. 25. Radioactivity of uranium-X

process of separating the active uranium-X, followed by its decay, can be repeated apparently indefinitely. The conclusion must consequently be drawn that uranium is continually producing the active product uranium-X, and that this gradually decays and loses its activity. Similar phenomena were shown by E. Rutherford and F. Soddy (1902) to occur when thorium was precipitated as hydroxide; in this case most of the activity was retained by the solution, the thorium hydroxide being almost inactive. In the course of time, however, the latter recovered the original activity of the thorium compound, but that of the residue obtained by evaporating the solution gradually declined.⁴

Radioactive Emanations.—An important observation made by E. Rutherford in 1900 was that thorium gave off spontaneously a radioactive

gas or **emanation**, which could be carried away in a current of air; shortly afterward, radium and actinium were found to evolve similar emanations. These were true gases, for they could be condensed by means of liquid air, but they were very inert chemically, and it was concluded that they were related to the inactive gases of the atmosphere. The emanations, particularly that from actinium, lost their activity relatively rapidly, but at the same time an "induced" activity appeared to be produced on solid bodies in contact with the gas. This so-called induced activity was later traced to solid radioactive material, called the **active deposit**, produced from the emanations. Each of the three emanations gave its characteristic active deposit, the activity of which, like that of other radioactive materials, decayed at a definite rate.

The Theory of Radioactive Disintegration.—With the object of correlating these remarkable observations E. Rutherford and F. Soddy (1903) proposed the revolutionary theory of **radioactive disintegration**. It was suggested that the atoms of radioelements, unlike those of stable elements, undergo spontaneous disintegration with the emission of α - or β -particles, and the formation of atoms of a *new* element which is quite different physically and chemically from its parent. This new element may in turn also be unstable and emit a particle with the production of still another element, and there may in fact be a succession of transformations each accompanied by its characteristic rays. The atomic weights and other properties of the elements produced in the radioactive disintegration depend on the nature of the particle emitted. For example, a radium atom, of mass 226 units on the atomic weight scale, gives off an α -particle, with a range of 3.39 cm. in air, having a mass of 4 units; the mass of the residual atom must, therefore, be 222 units. A new element of atomic weight 222 should thus be obtained, and this appears to be the radioactive emanation now known as **radon**. The density of the gas was actually measured with the aid of a sensitive microbalance, since only very minute amounts were available, and the atomic weight found to be 223 (W. Ramsay and R. Whytlaw-Gray, 1910). Radon is itself very radioactive, giving off an α -particle of range 4.12 cm., and yielding the solid radioelement radium-A which is the parent of the active deposit referred to above. This substance also disintegrates, and there are a number of further stages until a nonradioactive **end product** is reached. Every emission of an α -particle means that the new element has an atomic weight 4 units less than its parent, but the emission of a β -ray involves no appreciable change in atomic weight, because of the negligible mass of an electron. In spite of the apparent constancy of atomic weight, the chemical and physical properties of the product are quite different from those of its parent. Three radioactive disintegration series have been identified, viz., the uranium series, in which the element radium is included, the actinium series and the thorium series; they will be considered in detail later.

The results obtained in the chemical separation experiments can now be explained. The element uranium, for example, has a relatively feeble activity, but its disintegration product uranium-X, which is precipitated together with the iron on addition of ammonium carbonate, is very active. When the two elements are separated chemically, the feebly active uranium remains in solution whereas the highly active uranium-X is in the precipitate. In the course of time the uranium-X disintegrates further, the product being less active, and so there is a gradual decay in the activity of the precipitate. The uranium in solution, however, also continues to disintegrate and so produces more uranium-X, and hence the radioactivity increases until a certain equilibrium proportion, equal to that initially present, is attained. Actually, the so-called uranium-X is a mixture of two or three elements which disintegrate quite rapidly and so have a strong activity, but they soon yield a product, uranium-II, with a long life and feeble activity; for this reason the activity appears to fall almost to zero.⁵

Rate of Disintegration: Radioactive Constants.—If a radioactive substance, such as uranium-X, is separated from its parent and its activity is measured from time to time, the rate of decay is found to follow an exponential law; that is, at any instant the rate of decay is proportional to the activity at that instant. If I is the activity at any time t , then

$$-dI/dt = \lambda I, \quad (2)$$

where λ is a constant; upon integration this becomes

$$I = I_0 e^{-\lambda t}, \quad (3)$$

where I_0 is the initial activity at zero time. Since the activity may be assumed to be proportional to the number of atoms that have not yet disintegrated, it follows that

$$n = n_0 e^{-\lambda t}, \quad (4)$$

n_0 and n being the number of atoms of the particular radioelement, whose decay is being studied, initially and after time t , respectively.* The quantity λ is called the **radioactive constant**, and is characteristic of the element responsible for the activity being measured, irrespective of its physical condition or state of chemical combination.

The exponential law of decay implies that there is a definite probability of any particular atom disintegrating at a given moment, and this is proportional to the number of atoms present at that moment. The life of any radioactive atom, that is, the length of time it can exist before it disintegrates, can thus have all possible values from zero to infinity; in this way the gradual decay of activity is explained, for otherwise all the atoms would decay at the same instant. It is, nevertheless, useful to determine an **average life period** for an aggregate of a large number of atoms; it may be calculated as follows. Since the activity I is proportional to the number of atoms n which have not disintegrated, (2) may be written

$$-dn/dt = \lambda n \quad (5)$$

* It should be noted that the word *initially* does not refer to the instant at which the particular atoms were produced from their parent, but to the commencement of reckoning the time t ; the equation can, therefore, be applied over any convenient time interval.

and so, using the value of n from (4), it follows that

$$-dn = \lambda n_0 e^{-\lambda t} dt. \quad (6)$$

The number of atoms disintegrating in the interval between t and $t + dt$ is equal to dn , and since the interval dt is very small, dn may be taken to be the number disintegrating at the time t . If this time is reckoned from any instant, it is clear that dn , as given by (6), ignoring the sign, represents the number of atoms having an expectation of life t after that instant. The period of average life (τ) is obtained by multiplying every possible life period (t), from zero to infinity, by the number of atoms (dn) having that expectation of life, and then dividing the sum of these products by the total number of atoms (n_0) present at the beginning of the time; thus,

$$\begin{aligned} \tau &= \int_0^{\infty} \frac{t dn}{n_0} = \int_0^{\infty} t \lambda e^{-\lambda t} dt \\ &= 1/\lambda. \end{aligned} \quad (7)$$

The average life τ of a radioactive atom is thus equal to the reciprocal of its radioactive constant λ ; it applies at any time, and so is a measure of the *mean expectation of future life* of the atoms present at any instant.

The **half-life period** (T) is also an important quantity, for it gives the time which must elapse for the radioactivity to decay to half its value at any instant. It may be obtained by putting n equal to $\frac{1}{2}n_0$ in (4); thus,

$$\frac{1}{2}n_0 = n_0 e^{-\lambda T}; \quad \therefore T = \ln 2/\lambda = 0.693/\lambda. \quad (9)$$

In the time T the activity is reduced to one-half of its initial value, and so in the time nT the activity is decreased to $(\frac{1}{2})^n$; theoretically, therefore, the activity never falls to zero, although in ten times the half-life period it is reduced to 0.1 per cent of its value at the moment from which the time is reckoned.

The study of radioactive decay provides another means for calculating the Avogadro number. It has been recorded (p. 122) that 1 g. of radium gives off α -particles at the rate of 11.6×10^{17} per annum, and if each particle results from the disintegration of one atom it follows that $-dn/dt$ is 11.6×10^{17} , with the time in years, where n is the number of separate atoms in 1 g. The half-life period T of radium is known from decay and other measurements to be 1590 years, and so λ is $0.693/1590$ years; substitution in (5), therefore, gives the value of n , the number of atoms in 1 g. of radium. Since the atomic weight is 226, the quantity $226n$ is the number of atoms in 1 g. atom of radium, and this is the Avogadro number N ; hence,

$$\begin{aligned} N &= 226n = \frac{226 \times 11.6 \times 10^{17} \times 1590}{0.693} \\ &= 6.0 \times 10^{23}, \end{aligned}$$

in agreement with the results already obtained.

The Geiger-Nuttall Rule.—A relationship between the radioactive constant, or the period of average life, of a radioelement and the range of the α -particles emitted by it, was discovered experimentally by H. Geiger and J. M. Nuttall (1911). If $\log \lambda$ is plotted against the logarithm of the range, i.e., $\log R$, for a series of elements, an approximately straight line is obtained for each of the three radioactive series; thus,

$$\log \lambda = b + c \log R, \quad (10)$$

where b is somewhat different for the three series, but c , the slope of the lines, is practically the same in each case. It is seen, therefore, that the shorter the life of a radio-element the longer the range, and the greater the energy, of the α -particle it expels. The Geiger-Nuttall rule, as expressed by (10), is of great value, for by its means the life periods of very long-lived or very short-lived elements can be calculated from the range of the α -particles they emit. Since the range of an α -particle is related to its initial velocity (p. 122), viz., $v_0^3 = \sigma R$, it follows that

$$\log \lambda = b' + c' \log v_0, \quad (11)$$

thus giving a connection between the average life period of an element and the velocity of the α -particle expelled from it.

Wave Mechanics and Disintegration.—An equation, of which the Geiger-Nuttall rule is a first approximation, has been deduced from the wave mechanical theory of the nucleus (G. Gamow, 1928; R. W. Gurney and E. U. Condon, 1928). It is generally accepted that α -particles originate in the nucleus of a radioactive atom (cf. p. 136), and the curve representing the variation of potential energy of the particle with the distance from the interior of the nucleus is of the type shown in Fig. 26. From the scattering of α -particles by an atomic

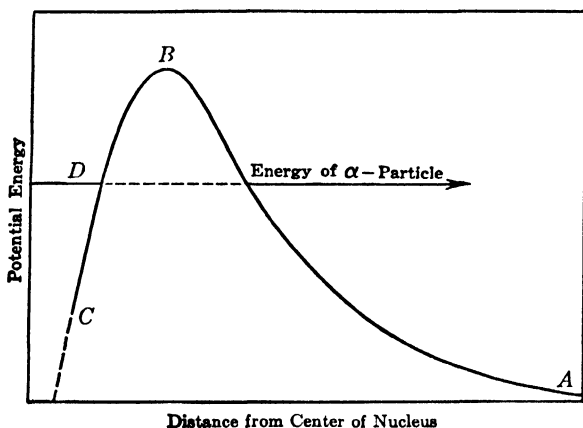


FIG. 26. Emission of α -particles from a nucleus (Gurney and Condon)

nucleus, it is known that as the nucleus is approached the potential energy of the particle is greatly increased; in other words, there is a strong force of repulsion between the α -particle and the atomic nucleus. This is shown by the rapid rise in the curve AB as the nucleus is approached. Since α -particles probably exist as such in many atomic nuclei, it is obvious that the curve AB cannot rise continuously, but must fall again, e.g., along BC , at distances very close to, and in the interior of, the nucleus. The distance from the center of the nucleus to the position of maximum potential energy B , represents the effective nuclear radius, which is of the order of 10^{-12} cm. (p. 24). If an α -particle is to leave the nucleus, as in a radioactive change, then according to classical ideas it must acquire sufficient energy to carry it over the barrier represented by the maximum of the potential energy curve. This is calculated to be equivalent to at least 14×10^{-6} erg per particle, but actually the fastest

α -particles have only about half this energy. The difficulty has been overcome however, by means of wave mechanics. The association of a wave function with every particle (p. 75) means that there is a definite, even if small, probability of the passage of a particle from one region to another at the same energy, even if there is a considerable potential barrier in between. If the energy of an emitted α -particle is shown by the arrow in Fig. 26, then an α -particle in the nucleus having this same amount of energy, as represented by the line at D , will have a definite probability of escape. It is to be expected that the greater the energy the greater will be this probability, so that an α -particle with high energy, and hence with high velocity, is more likely to be able to leave the nucleus than one with low energy and velocity. A high probability of escape of α -particles is equivalent to saying that the radioactive decay constant is large, and so the qualitative interpretation of (11) is evident. By the use of the appropriate wave equations it is shown to be quantitatively correct also, provided the range of velocities is not very great.

Radioactive Equilibrium.—The rate of decay of a given radioelement must also represent the rate at which its disintegration product is being formed. The latter will itself disintegrate at a rate dependent on the amount present; this will be small initially, but will increase with time. It is apparent that a state of equilibrium will eventually be reached, in any given series, when the rate of formation of any element from its parent is equal to that at which it is itself disintegrating. This may be expressed mathematically, thus

$$\frac{dn_1}{dt} = \frac{dn_2}{dt} = \frac{dn_3}{dt} = \text{etc.}, \quad (12)$$

where n_1, n_2, n_3 , etc., represent the numbers of atoms of different radioelements in the series present at equilibrium; from (5) it then follows that

$$\lambda_1 n_1 = \lambda_2 n_2 = \lambda_3 n_3 = \text{etc.} \quad (13)$$

At radioactive equilibrium, therefore, the amounts of different radioelements present are inversely proportional to their decay constants, or directly proportional to their half-life or average-life periods. The radioactive constant of an element of long life can thus be determined from a knowledge of the ratio of the amount present at equilibrium to that of an element of directly measurable rate of decay.

The time taken for complete equilibrium to be established depends on the life periods of the elements involved; the longer the average life of any element the more slowly is equilibrium with its products reached. It is because uranium-II has a half-life of about a million years that the activity of its parent, uranium-X, decays almost to zero in a short time (see Fig. 25), although several of the successive stages are highly active. It is only in minerals, which have been in the earth for millions of years, that true radioactive equilibrium exists. All uranium minerals, for example, contain 3.3×10^{-7} part by weight of radium to one part of uranium, a fact which not only indicates the attainment of equilibrium but also shows that the element radium represents a stage in the disintegration of uranium.

Another important consequence of the principle of radioactive equilibrium is that, in spite of their very different amounts, all elements present at equilibrium are decaying at the same rate, i.e., the same number of atoms disintegrate in a given time; this will be evident from an examination of (12) and (13). Radium itself emits an α -particle, and among its products are three

other elements giving off α -rays, viz., radon, radium-A and radium-C', with which it comes into equilibrium rapidly; * hence starting with radium, free from its parents, and allowing virtual, although not complete, equilibrium to be attained the resultant rate of emission of α -particles will be four times that for radium alone. This fact is utilized, for example, when determining the rate of production of helium from radium; one quarter of the total volume is obtained from α -particles emitted by the radium itself.

Separation of Radioelements.—For the purpose of studying the individual radioelements it is necessary to be able to make at least partial separations, and various methods have been applied in different cases. Chemical processes, some of which have been already mentioned, are often employed, as for example, the solubility of uranium compounds in ammonium carbonate and the precipitation of radium as sulfate. Polonium is precipitated from an acid medium by hydrogen sulfide, and ionium and thorium are thrown out of solution as hydroxides on the addition of ammonia. Since most radioelements are present only in very minute amounts, the separation can only be carried out if an appreciable quantity of an element is added which is itself precipitated at the same time, e.g., barium with radium, and bismuth with polonium. Electrochemical methods, depending on the displacement of one metal by another, e.g., radium-C by nickel, or on deposition using graded potentials, have also been employed for the separation of radioelements.

The production of a gaseous emanation in each radioactive series provides a convenient means of separating all the elements preceding it; these emanations soon disintegrate giving the solid active deposits which can be collected on a negatively charged wire inserted in the gas. This forms a useful method of concentrating radioactive material. The active deposits also decay rapidly, and in each case elements designated by the letters A, B and C are formed in the first three stages. By heating the wire, the B-elements, e.g., radium-B, are readily volatilized whereas the C-elements, e.g., radium-C, remain. Other separations in the active deposit may be made by utilizing the phenomenon of recoil. When an α -particle of mass 4 is expelled from a radioactive atom of mass about 220, the principle of the conservation of momentum requires the resultant atom to recoil with a velocity of $4/216$, i.e., $1/54$, of that of the α -particle, that is, at about 3 to 4×10^7 cm. per sec. This relatively high speed permits the atoms to leave the plate or wire on which their parent was deposited, provided the latter is in the form of a thin layer; in an evacuated space the recoil atoms can travel for some distance and be collected on a negatively charged plate or wire. In this way radium-D can be separated from the radium active deposit.⁶

Disintegration Series.—As a result of a great deal of investigation, involving both physical and chemical methods, such as determination of atomic weights when possible, the analysis of abnormal decay curves, which show the presence of several elements with life periods of the

* Since a long-lived product (RaD) intervenes, the concentration of RaF, which also gives off α -rays, does not reach equilibrium rapidly. The α -ray activity of RaC, viz., 0.04%, is so small that it may be neglected (see Table 17).

same order, measurement of the ranges of particles emitted, and so on, it has been possible to characterize most of the members of the three main radioactive series. The results obtained are summarized in Tables 17, 18 and 19, below. The loss of an α -particle involves a decrease of four units in the atomic weight, but there is no change when a β -particle is emitted; the figures in parentheses have been estimated in this manner. The method of arriving at the atomic number will be explained shortly. Attention may be called to the **branched disintegration** which sometimes occurs, that is, when the same parent, e.g., radium-C, gives rise to two entirely different products, viz., radium-C' and -C''. In this case, and in some others, the two branches eventually yield the same radioelement, e.g., radium-D.

In addition to the radioelements mentioned in Tables 17, 18 and 19, brief reference must be made to others whose position is uncertain. The β -active uranium-Z was believed to be the result of a branched β -ray disintegration, to a very small extent, of uranium-X₁, but this is not certain; an alternative possibility is that it arises from "internal conversion" (p. 171) of a γ -ray from uranium-X₂. In any event, uranium-Z presumably emits a β -particle and becomes uranium-II, the half-life being 6.7 hr. It was thought, at one time, that uranium-Z formed uranium-Y upon disintegration, and this was the parent of protoactinium. It is more probable, however, that the parent of the actinium series is an element of atomic number 92 and atomic weight 235, called actinouranium (AcU);

TABLE 17. THE URANIUM SERIES

Element	Symbol	At. Wt.	At. No.	Rays	Half-life Period
Uranium-I	U _I	238.2	92	α	4.5×10^9 yr.
↓					
Uranium-X ₁	U _{X1}	(234)	90	β	24.5 days
↓					
Uranium-X ₂ *	UX ₂	(234)	91	β	1.14 min.
↓					
Uranium-II	U _{II}	(234)	92	α	2.7×10^5 yr.
↓					
Ionium	Io	(230)	90	α	8.3×10^4 yr.
↓					
Radium	Ra	226.05	88	α	1590 yr.
↓					
Radon (Emanation)	Rn	222	86	α	3.82 days
↓					
Radium-A	Ra _A	(218)	84	α	3.05 min.
↓					
Radium-B	Ra _B	(214)	82	β	26.8 min.
↓					
Radium C ₁ 99.96%	Ra _{C1}	(214)	83	β and α	19.7 min.
↓					
Radium-C''	Ra _{C''}	(214)	84	α	1.5×10^{-4} sec.
↓					
Radium-C'	Ra _{C'}	(210)	81	β	1.32 min.
↓					
Radium-D	Ra _D	(210)	82	β	22 yr.
↓					
Radium-E	Ra _E	(210)	83	β	50 days
↓					
Radium-F (Polonium)	Ra _F	(210)	84	α	140 days
↓					
Radium-G (Uranium lead)	Ra _G	206	82	—	—

* Possibly undergoes a branched transition, forming Uranium-Z, by internal conversion; this disintegrates as follows: UZ (234, 91, β , 6.7 hr.) \rightarrow Uranium-II.

TABLE 18. THE ACTINIUM SERIES

Element	Symbol	At. Wt.	At. No.	Rays	Half-life Period
Protoactinium*	Pa	(231)	91	α	3.2×10^4 yr.
Actinium	Ac	(227)	89	β	13.5 yr.
Radioactinium	RaAc	(227)	90	α	18.9 days
Actinium-X	AcX	(223)	88	α	11.2 days
Actinon (Emanation)	An	(219)	86	α	3.92 sec.
Actinium-A	AcA	(215)	84	α	1.83×10^{-3} sec.
Actinium-B	AcB	(211)	82	β	36 min.
Actinium-C	AcC	(211)	83	α and β	2.16 min.
99.84% 0.16% Actin.-C' Actin.-C''	AcC' AcC''	(211) (207)	84 81	α β	5×10^{-3} sec. 4.76 min.
Actinium-D	AcD	(207)	82	—	—

* Possibly preceded by Actinouranium (AcU, 235, 92, α , 7.1×10^8 yr.) \rightarrow Uranium-Y (UY, 231, 90, β , 24.6 hr.) \rightarrow Protoactinium.

TABLE 19. THE THORIUM SERIES

Element	Symbol	At. Wt.	At. No.	Rays	Half-life Period
Thorium	Th	232.12	90	α	1.39×10^{10} yr.
Mesothorium 1	MsTh ₁	(228)	88	β	6.7 yr.
Mesothorium 2	MsTh ₂	(228)	89	β	6.13 hr.
Radiothorium	RaTh	(228)	90	α	1.90 yr.
Thorium-X	ThX	(224)	88	α	3.64 days
Thoron (Emanation)	Tn	(220)	86	α	54.5 sec.
Thorium-A	ThA	(216)	84	α	0.16 sec.
Thorium-B	ThB	(212)	82	β	10.6 hr.
Thorium-C	ThC	(212)	83	β and α	60.5 min.
65% 35% Thorium-C' Thorium-C''	ThC' ThC''	(212) (208)	84 81	α β	3×10^{-7} sec. 3.1 min.
Thorium-D	ThD	208	82	—	—

it is suggested that this emits an α -particle, yielding uranium-Y, and the latter gives off a β -particle and forms protoactinium.

Radioelements and the Periodic Table.—The discovery of some forty elements with different radioactive properties, having atomic weights between 206 and 238, raised the question of their arrangement in the periodic classification of the elements. The position was partly simplified when it was found that several radioelements possessed identical chemical and physical properties, and differed only in the nature of the disintegration they suffered. A simple example already encountered is the similarity of the three emanations, all inert gases like those of the zero group of the periodic table. In the course of a systematic investigation, made by A. Fleck (1913), under the direction of F. Soddy, certain groups of

elements were found to be quite inseparable by chemical methods, so that they could be regarded as being identical chemically. This identity frequently extended to nonradioactive elements, as the results given below indicate; the elements in each horizontal line are chemically inseparable.

Radioactinium and Thorium.

Radium-*B* and -*D*, Thorium-*B*, Actinium-*B* and Lead.

Radium-*C* and -*E*, Thorium-*C*, Actinium-*C* and Bismuth.

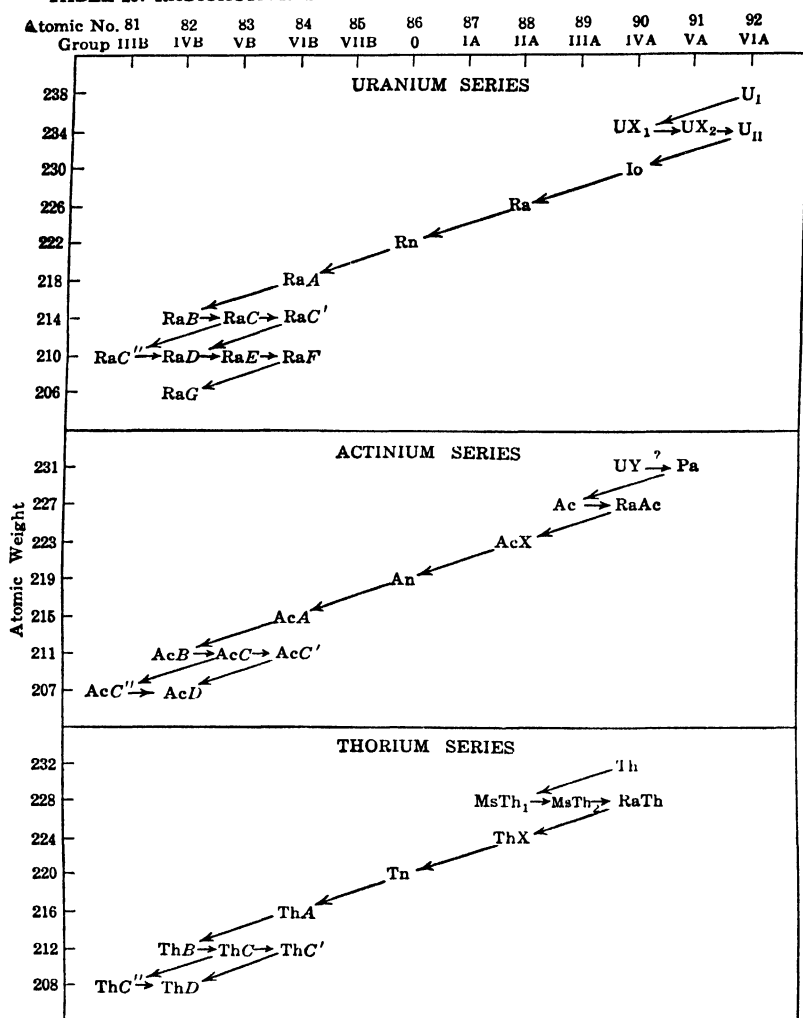
Uranium-*I* and Uranium-*II*.

In 1911 F. Soddy had suggested that when a given radioelement expels an α -particle the product occupies a position two places lower than its parent in the periodic table, and two years later (1913) a wider generalization was proposed independently by K. Fajans, A. S. Russell, and F. Soddy, which has become known as the **group displacement law**. It coordinates the results described above and accounts for that fact that the large number of radioelements can be fitted into the limited number of spaces available in the periodic system. The law may be stated in the following form: *when an α -particle is emitted in a radioactive change, the product is displaced two places to the left in the periodic table, that is, the atomic number is two less than that of its parent, but the emission of a β -particle results in a displacement of one place to the right, the atomic number of the product being now one greater than that of its parent.* The successive emission of an α - and two β -particles thus brings an element back to its original position in the table, although its atomic weight is four units less.

The essential facts of the three disintegration series are given in Table 20 in accordance with the group displacement law; an α -ray change is indicated by an arrow sloping downward, implying a decrease of four in atomic weight, and two places to the left, to represent the decrease of two in the atomic number. The emission of a β -particle is shown by a horizontal arrow, since there is no change in atomic weight, pointing one place to the right because the atomic number is increased by one. In this manner the changes in both atomic weight and atomic number are represented simultaneously.⁷

Isotopes.—A comparison of Table 20 with the examples given above of elements which are inseparable chemically, shows that each group is to be found in one column of the table, so that they all have the same atomic number; it appears, therefore, that elements with the same atomic number have identical chemical properties, although their radioactive properties may be different. F. Soddy (1913) proposed the name **isotopes** (Greek: *same place*) for such elements because they occupy the same place in the periodic table; thus, it can be seen that uranium-*I* and uranium-*II* are isotopic with each other, and the *C*'-elements of the three series are isotopes of the nonradioactive element bismuth, atomic number 83. The *C*''-elements in the column representing the atomic number 81 are all isotopes of thallium, whereas the seven elements in the next column, atomic number 82, are isotopic with lead.

TABLE 20. RADIOACTIVE SERIES AND THE GROUP DISPLACEMENT LAW



The Nucleus and the Group Displacement Law.—Before proceeding to consider other aspects of the subject of isotopes it is of interest to examine the group displacement law from the standpoint of the nuclear atom. It is probable that in radioactivity the α - and β -particles are ejected from the nucleus of the atom; this is in harmony with the observation that their rate of emission is independent of the state of combination of the atom or of the physical conditions, even extremes of temperature having no effect. The loss of an α -particle, carrying two positive charges, from the nucleus means that the charge of the latter will be decreased by two, and so the resultant nucleus will correspond to that of an element with atomic number two units less than that of the original

element. The expulsion of a β -particle involves first the conversion of a neutron into a proton and an electron, perhaps together with a neutrino, so that when the electron is expelled the positive charge of the nucleus must increase by unity; the atomic number of the product should consequently be one unit greater than that of the parent. It will be noted that in the former case the resultant atom will have two extranuclear electrons too many for electrical neutrality, whereas in the latter there will be one too few. This excess or shortage of electrons outside the nucleus is easily adjusted by exchange with surrounding bodies; eventually the α -particle captures two electrons to become a neutral helium atom, whereas the β -particle finds its way into the outer electronic system of one atom or another.

Consideration of the changes undergone by the nucleus throws some light on the problem of isotopes. It has been mentioned that when an element loses one α - and two β -particles, in successive stages, the product occupies the same place in the periodic table as, and is consequently isotopic with, the original element. If an α -particle may be regarded as made up of two protons and two neutrons (p. 30), the changes in the constituents of the nucleus may be represented thus:

	Protons	Neutrons
Loss of α -particle	- 2	- 2
Loss of two β -particles	+ 2	- 2

The net result is, therefore, the loss of four neutrons from the nucleus, but the number of protons, and hence the charge, remains unchanged. The difference between isotopes is, therefore, that their nuclei contain different numbers of neutrons, although the same number of protons; the consequence is that the elements have the same atomic numbers, but different atomic weights, for the number of protons determines the former, and the total number of neutrons and protons governs the latter. In the type of isotopy just considered the atomic weights differ by four units, but this is not necessarily always the case; it is seen from Table 20 that the isotopes of lead, for example, have atomic weights of 206, 207, 208, 210, 211, 212 and 214. The nuclei all contain 82 protons, for this is their atomic number, but the number of neutrons is different in each case, being 82 less than the atomic weight. The corresponding differences in nuclear structure lead to the variations in radioactive properties.

Nonradioactive Isotopes.—It will be seen from Tables 17 and 18 that the last elements, that is the presumably non-radioactive end-products, of the uranium and thorium series are radium-*G* and thorium-*D* respectively; these both have the atomic number 82 and are isotopes of lead, although their atomic weights are 206 and 208, respectively, as compared with 207.2 for ordinary lead. Uranium and thorium minerals in nature are almost invariably associated with lead, and if some, at least, of the latter were of radioactive origin, having been formed by gradual disintegration from a parent element, the atomic weight should differ from that of lead obtained from a source not associated with radioactive material. The inevitable, if surprising, conclusion to be drawn from the group displacement law is that the atomic weight of lead should vary from one source to another. Preliminary determinations by F. Soddy and H. Hyman (1914) showed that the atomic weight of lead derived from

the mineral Ceylon thorite was greater than that of ordinary lead, and shortly afterwards T. W. Richards and M. E. Lambert (1914) found a much lower value for lead from the uranium minerals pitchblende and uraninite. Further estimations by the most experienced workers in the field of atomic weights have confirmed these conclusions; some of the recent data are summarized in Table 21. It may be emphasized that in spite of the differences in atomic weight, the various specimens of lead and their salts were identical chemically, and were quite pure, in the ordinary sense. Corresponding to the difference of atomic weight, a

TABLE 21. ATOMIC WEIGHT OF LEAD FROM RADIOACTIVE SOURCES

Origin	Atomic Wt.	Investigators
Uranium Minerals		
Cleveite, Norway	206.08	T. W. Richards and C. Wadsworth, 1916
Bröggerite, Norway	206.01	T. W. Richards and C. Wadsworth, 1916
Pitchblende, Katanga	206.35	O. Hönigschmid and L. Buckenbach, 1923
Pitchblende, Katanga	206.35	G. P. Baxter, <i>et al.</i> , 1937
Kolm, Sweden	206.01	G. P. Baxter and A. D. Bliss, 1930
Thorium Minerals		
Thorite, Ceylon	207.77	O. Hönigschmid, 1917
Thorite, Norway	207.9	O. Hönigschmid, 1919

difference of density was observed, although the atomic volume is approximately constant for the various forms of lead. The characteristic X-ray spectra also appear to be identical, as is to be expected according to Moseley's law for elements having the same atomic number. In accordance with theoretical requirements, the positions of corresponding lines in the optical spectra of the isotopes of lead are slightly displaced with respect to one another.

The results recorded above not only provide strong confirmation of the theories of radioactivity, they also prove the existence of isotopes as different forms of the same element, differing in atomic weight, but having the same nuclear charge and occupying the same position in the periodic table, and so having identical chemical properties. The existence of radioactive isotopes, such as the seven elements of atomic number 84, may perhaps not be considered surprising, but the experiments with lead showed that even nonradioactive elements could exist in isotopic forms. This discovery led to the search for other isotopic elements, and in this connection the most valuable results have come from a further study of the positive and anode rays already described (p. 2).⁸

NONRADIOACTIVE ISOTOPES

Positive Ray Analysis.—A thin beam of positive rays will normally move in a straight line, but if subjected to the action of an electric or magnetic field, the beam will be deviated from its original direction. Imagine a single positively charged particle moving downward perpendicular to the plane of the paper, and striking it at the point *O* in Fig. 27;

if a uniform electric field of strength X is now applied so that the particle is deflected to the right, it will strike the paper at E , the distance x from O to E being given by

$$x = k_1 \frac{Xe}{mv^2}, \quad (14)$$

where k_1 is a constant depending on the dimensions of the apparatus, e is the charge on the particle which must be a multiple of the electronic charge, and m and v are the mass and velocity, respectively, of the positive particle. If, on the other hand, instead of an electric field, a magnetic field H is applied to deflect the particle upward, so that it strikes at M , the displacement OM is given by

$$y = k_2 \frac{He}{mv}, \quad (15)$$

where k_2 is also dependent on the apparatus. By applying the electric and magnetic fields simultaneously the particle is deflected to P , the coordinates of which are given by x and y of (14) and (15); if v is eliminated from them it follows that

$$\frac{y^2}{x} = k \frac{H^2}{X} \cdot \frac{e}{m}, \quad (16)$$

where k is another constant related to k_1 and k_2 . If instead of one particle a beam of particles is considered, all having the same charge to mass ratio, that is, e/m is constant throughout, but varying in velocity, then it follows, from (16), that provided the strengths of the electric and magnetic fields are fixed,

$$y^2/x = \text{constant}. \quad (17)$$

This is the equation for a parabola, and so the particles, instead of falling on one spot P , will fall on a series of points whose coordinates are given by (17); that is to say, the points will constitute a parabolic curve, such as AA' in Fig. 27. An examination of (14) and (15) shows that the displacements x and y depend on the velocity of the particle, the greater the velocity the smaller the displacement, and *vice versa*; the different points making up the parabolic curve represent, therefore, particles with different velocities. The fast moving particles are deflected only to a small extent, whereas the slow particles are considerably deflected. The continuity of the curve AA' implies that particles having all possible velocities between certain limits are present in the positive rays.

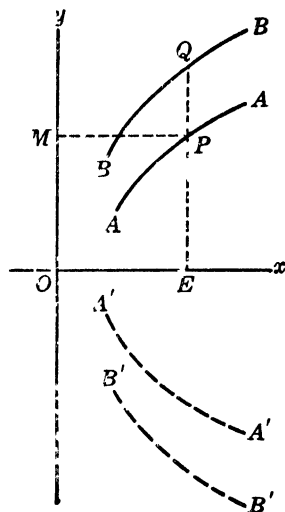


FIG. 27. Deflection of positive rays

A beam of particles for which e/m is different from that already considered will fall on another parabolic curve BB , and from (16) it is seen that the greater the charge to mass ratio the greater will be y^2/x , and so the higher the position of the curve. In Fig. 27, therefore, BB would be formed by particles having a larger value of e/m than for AA ; if the ordinates y_A and y_B of points on the two curves, represented by the distances PE and QE , correspond to a constant value of the abscissa x , then it is seen from (16) that

$$\left(\frac{y_A}{y_B}\right)^2 = \frac{(e/m)_A}{(e/m)_B} \quad (18)$$

and if the charges carried by the particles may be assumed to be equal, it follows that

$$\left(\frac{y_A}{y_B}\right)^2 = \frac{m_B}{m_A}. \quad (19)$$

It should thus be possible to compare the masses of positively charged particles, such as are present in positive rays, by observing their deflections in simultaneously applied electric and magnetic fields. If a beam of positive rays containing particles of different masses is employed then they will be sorted out in such a manner that all having the same mass, or more strictly the same e/m value, will fall on one parabolic curve; a method is thus available for the analysis of the constituents of positive rays.

Positive Ray Parabolas.—The first application of the principle of positive ray analysis was made by J. J. Thomson (1912). The rays were produced in a discharge tube at a suitable low pressure; the pierced aluminum cathode was connected to a fine-bore brass tube through which the positive rays streamed as a fine beam. They were then made to pass between electric and magnetic fields applied simultaneously, and finally allowed to fall on to a photographic plate. On development, the latter showed a series of parabolic streaks (Fig. 28), each corresponding to a definite ratio of charge to mass of the particles present in the positive rays, as postulated above. In order to determine the coordinates for the various curves, so that the e/m values might be compared, it was necessary to fix the origin and the x -axis. The position of the former is easily obtained from the spot produced by the undeflected beam, and the latter is determined by keeping the electric field in the same direction and reversing the magnetic field; the parabolic curves $A'A'$ and $B'B'$ (Fig. 27), which are reflections of AA and BB in the x -axis, appear. By reversing the electric field another set of curves may be obtained on the same plate, as seen in the upper left-hand portion of Fig. 28. From (19) it is evident that the determination of the mass of the particles producing any particular curve necessitates a comparison of the ordinate with that

of a standard substance of known mass; as a general rule sufficient amounts of oxygen or carbon dioxide will be present to give the necessary

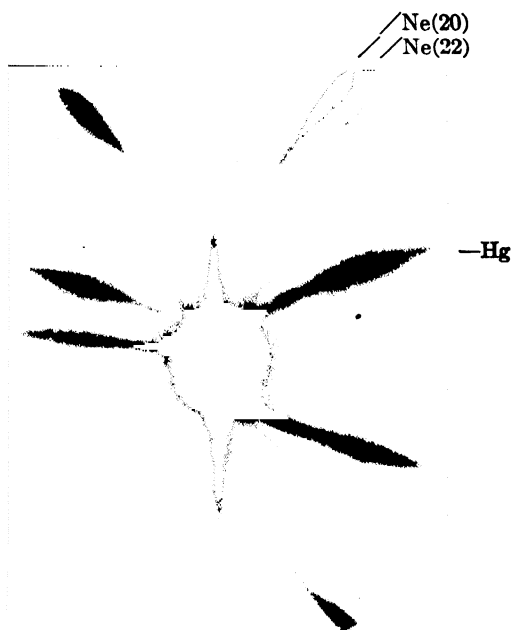


FIG. 28. Positive ray parabolas (J. J. Thomson)

comparison lines, but if not, traces may be deliberately introduced into the discharge tube. Carbon dioxide, for example, gives lines produced by the following particles: CO_2^+ , CO^+ , C^+ , O^+ , C^{++} and O^{++} .⁹

Positive Rays and Isotopes.—Although the positive ray photographs showed a number of interesting features, such as curves corresponding to particles which have no normal stable existence, e.g., CH_3 , CH_2 , CH from methane, and which must evidently exist for short periods in a discharge tube, the discussion here will be restricted to the connection between positive rays and isotopes. An examination of the parabolas obtained with neon in the discharge tube (Fig. 28) shows the presence of a faint line due either to a single charged particle of mass 22 or to a particle of mass 44 carrying a double charge. The latter might possibly have been produced by carbon dioxide molecules, but this was disproved by the fact that cooling the neon in liquid air, which would have condensed all the carbon dioxide, had no effect on the intensity of the 22-line. Further, it is a matter of experience that double charged molecular, as distinct from atomic, ions are rarely found, so that it is unlikely that CO_2^{++} particles/

would be present. In discussing the observations, J. J. Thomson (1913) considered the possibility that the 22-line might be produced by a compound of neon and hydrogen, NeH_2 , but in view of the known properties of neon this did not seem probable. It was evident, however, that ordinary neon gas contains a substance of relative mass 22, and an attempt to concentrate it was made by F. W. Aston (1913) by repeated fractional distillation of pure neon held in charcoal cooled in liquid air, and also by fractional diffusion at low pressures through clay pipe-stems. No positive results were obtained from the first method, but by means of the second, two extreme fractions having molecular weights, determined by density methods, of 20.15 and 20.28 were separated. This difference although small was definite so that it was regarded as proved that neon was not a simple substance; the two specimens, however, had apparently identical spectra, so that they contained only neon. When these experiments were made, the theory of isotopes was only just being developed, but by 1919, when Aston resumed the work, the concept of isotopes was established and it appeared probable that the 22-line found by Thomson was due to an isotope of neon. Using an improved apparatus, Aston obtained two positive ray parabolas with pure neon, corresponding to particles of relative masses 20.00 and 22.00; the ratio of the amounts, estimated from the intensities of the lines, was about 10 to 1, so that the average mass is 20.2, in excellent agreement with the known atomic weight of neon, viz, 20.18. It seems clear, therefore, that ordinary neon gas consists of a mixture of two isotopes having different atomic weights.

The positive ray apparatus for the study of isotopes was called by Aston the **mass spectrograph**, for it gives a series of lines determined by the masses of the particles present in the positive ray beam. It has been modified and improved in many ways so that it can now be used for the detection of isotopes, for estimating their relative amounts, and for determining their masses with an accuracy at least as great as that of the very best atomic weight measurements.

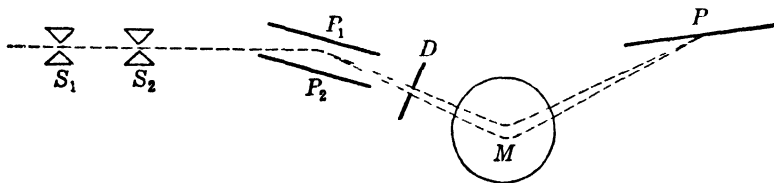


Fig. 29. Aston's mass spectrograph

Aston's Mass Spectrograph.—In order to increase the precision of the method it is necessary to obtain lines which are intense yet sharp; F. W. Aston (1919) achieved this end by the arrangement shown diagrammatically in Fig. 29. In Thomson's positive ray apparatus the magnetic and electric fields produced deflections in planes at right angles, but in Aston's

arrangement the deflections were in the same plane, but in opposite directions. The positively charged particles first pass through two very narrow parallel slits, S_1 and S_2 , and the resulting fine beam is spread out by means of an electric field applied across the plates, P_1 and P_2 . A section of the beam emerges through a small aperture in D and then passes between the pole pieces M of a large electromagnet; the magnetic field is so arranged as to bend the rays in the opposite direction to, although in the same plane as, that of the electric field. If the disposition of the apparatus is properly made the beam is brought to a focus on the photographic plate P , and on this a sharp line is produced for each type of particle present in the original beam. On development there is obtained, therefore, a **mass spectrum**, on which each line represents a definite m/e value (see Fig. 31). The range of the spectrum brought on the plate can be varied by altering the strengths of the deflecting fields, so that any particular masses can be studied at will. The determination of the mass of the particles producing a given line is made by comparison with that given by a particle of known mass; this can be done in several ways. In the "coincidence" method the strength of the electric field is altered until the position of the known line corresponds exactly to that originally occupied by the particles of unknown mass; the ratio of the two masses is then inversely proportional to that of the two field strengths [equation (14)]. For particles of small mass the "bracketing" method has been employed; it takes advantage of the fact that several of the lowest atomic masses are approximately one-half known masses, e.g., the m/e value for O^{++} is 8 000 (standard) for He^+ it is approximately 4, and for H_2^+ it is about 2. A photograph is taken with a constant magnetic field and a known potential, e.g., 250 volts, producing the electric field; the procedure is then repeated with potentials of $500 + 12$ and $500 - 12$ volts respectively. This has the effect of bracketing the line of unknown m/e by two lines, one on each side, due to particles having accurately known m/e values, approximately twice that of the particles being studied; from the relative positions of the three lines the unknown mass corresponding to the central line can be calculated.

For the investigation of elements forming volatile compounds the rays of positively charged particles were ordinary positive, i.e., canal, rays produced in a gas discharge tube into which the vapor was permitted to leak. When metallic elements were studied, however, use was made of anode rays (p. 2) a paste of graphite and a salt of the element being pressed into a hole drilled in the anode. The positive particles generated in this manner were accelerated by passing through a potential drop of about 20,000 volts.

In a later mass spectrograph built in 1925, F. W. Aston made several improvements in design and technique, although the main principles of the experimental method were unchanged. Some modifications were also made in the method for determining the mass values of the lines in

the spectrum, and it was claimed that the masses of particles could be estimated with an accuracy of 1 part in 10,000.¹⁰

Dempster's Method: Direction Focusing.—A different principle has been used by A. J. Dempster (1918–22) to focus beams of positive ions and to analyze them. The rays are produced either by heating salts of the metallic element being investigated or by bombarding them with electrons of low velocity; alternatively the metal itself may be vaporized by heating it electrically and the vapor ionized by electron bombardment. The positive particles thus obtained have very low velocities, and if they are then accelerated by passage through an electric field, the potential being about 1000 volts or more, they may be regarded as emerging from the field with uniform kinetic energy. If V is the accelerating potential and e the charge on the particle, the energy will be Ve , and this, being all kinetic, may be equated to $\frac{1}{2}mv^2$; thus

$$Ve = \frac{1}{2}mv^2, \quad (20)$$

where m and v are the mass and velocity, respectively, of the positive particle. A thin beam of ions accelerated in this manner passes through a narrow slit into a magnetic field between two semicircular iron plates; this causes the particles to move in a semicircular path, thus passing through a second slit, and on to a plate connected to an electrometer or similar device for measuring the ion current. It has been shown (p. 3) that the radius of curvature r of the path of a charged particle moving in a magnetic field H is determined by the relationship

$$e/m = v/Hr, \quad (21)$$

and so it follows, from (20), that for the positively charged particles

$$\frac{e}{m} = \frac{2V}{H^2r^2}. \quad (22)$$

In the apparatus under consideration only particles having a definite value of r can pass through the second slit and reach the plate and so be registered by the electrometer. The e/m values for the particular particles having a path of this radius is determined by the accelerating potential V and the strength of the magnetic field H . In practice, the latter is kept constant and the former varied steadily and the corresponding ion current measured for each value of V ; the results are plotted on a graph when a series of maxima are obtained. Each maximum corresponds to a set of particles having a definite charge to mass ratio; from a knowledge of the potential at which this occurs it is possible to calculate the value of e/m , from (22), provided H and r are known. These need not be measured directly, for the apparatus may be standardized with a substance of known mass. The relative amounts of the various particles can be estimated from the magnitude of the ion currents at the corresponding maxima. One advantage of Dempster's method is that it gives "direction focusing"; for ions having different initial directions, within limits, are brought to the same point by the 180° magnetic field, and in this way the intensity of the spectrum is increased.¹¹

Bainbridge's Method: Velocity Focusing.—In this method (K. T. Bainbridge, 1932), which gives results of a high order of accuracy, a ray of positive

particles obtained from a discharge tube passes through the slits S_1 and S_2 (Fig. 30) and then between the plates PP where it is subjected simultaneously to the action of opposed electric and magnetic fields. Only those particles remain undeflected, and so penetrate the slit S_3 , for which the displacements produced by the two fields exactly compensate; this will occur when Hev is equal to Xe (see p. 3), where H and X are the strengths of the magnetic and electric fields, respectively, e is the charge and v the velocity of the particle. It follows, therefore, that for all particles emerging from S_3 ,

$$Hev = Xe; \quad \therefore v = X/H, \quad (23)$$

and since the field strengths are constant, they will all have the same velocity

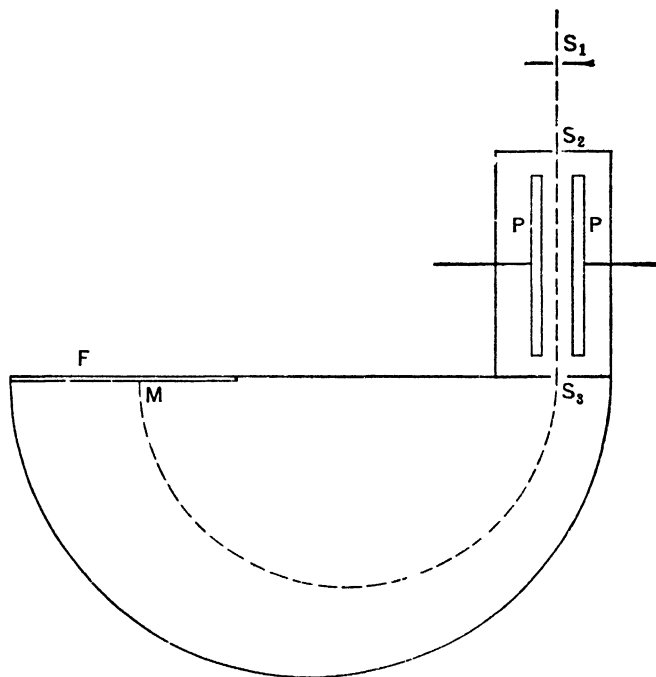


FIG. 30. Bainbridge's mass spectrograph

After leaving the slit S_3 these particles, all with equal velocities, enter another magnetic field, of strength H' , so that they are deflected into a semicircular path and strike the photographic plate F . From (21) it is seen that the radius of this path is equal to mv/cH' , and since v , e and H' are constant, it is directly proportional to the mass. The distance from S_3 to M is thus proportional to the mass of the particles striking the plate at M ; each set of particles having a definite value of m , or more correctly m/e , will produce a line on the plate. The linear relationship between the mass of the particle and the position of the line simplifies very considerably the accurate determination of the mass values which can be estimated to 1 part in 10,000.



FIG. 31. Mass spectra of cadmium, palladium and barium (Dempster)

Double Focusing Mass Spectrographs.—The more recent forms of mass spectrograph (A. J. Dempster, 1935; K. T. Bainbridge and E. B. Jordan, 1936; J. Mattauch and R. Herzog, 1936) employ both direction and velocity focusing, thus giving high intensity and a linear mass scale, as well as good resolving power so that the lines are well spaced. In one method the positive ions are passed through a radial electric field, the plates being bent into an arc instead of being flat, and then between the poles of a magnet arranged so that in conjunction with the electric field it gives velocity focusing. In this arrangement ions of varying initial directions are focused at the same time, so that the advantages of both schemes are obtained. Some mass spectra obtained by Dempster are shown in Fig. 31.

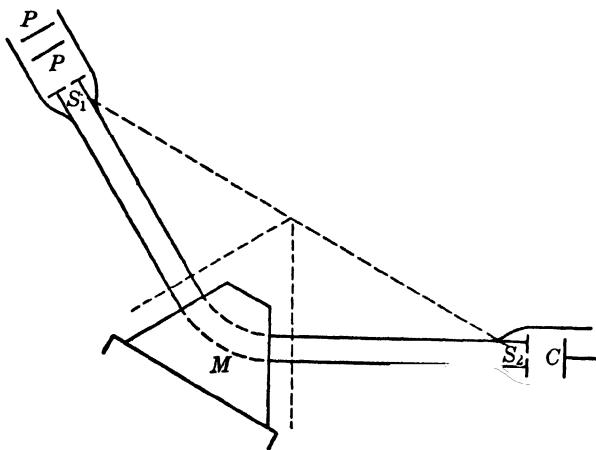


FIG. 32. Nier's mass spectrograph

A relatively simple but accurate form of mass spectrograph, employing the double focus principle, has been designed by A. O. Nier (1940); it is particularly valuable for the determination of abundance ratios, and it has been

used for analytical work. The positive ions, accelerated by passage through an electric field of known potential V , applied between the plates PP , pass through a slit S_1 (Fig. 32). They continue down a copper tube and are deflected through an angle of 60° by a magnetic field, strength H , between two V-shaped poles of the magnet M . The deflected beam, after passing through the slit S_2 , comes to a focus on the collector C , which is attached to a suitable amplification device permitting the strength of the ion current to be measured. The apparatus is arranged so that the charged particles, all of which have the same momentum, enter and leave the magnetic field in directions perpendicular to the pole faces of the magnet; the ion source, the apex of the V-shaped field, and the point of focus are then colinear. This is indicated by the broken line in the figure. The radius of curvature r of the ion path, which is constant, is given by (22); since the magnetic field H is held constant, ions of different e/m values can be brought to a focus at C by varying the potential V . The latter is obviously proportional to e/m , and so gives the mass of the particles, while their abundance is determined by the magnitude of the ion current.¹³

Detection of Isotopes by Optical Spectra.—A few isotopes, notably those of hydrogen, oxygen, nitrogen and carbon, were first detected by spectroscopic methods to which brief reference will be made. After evaporation of about 4 liters of liquid hydrogen to about 1 cc., the Balmer spectrum of the residue showed a faint line which could not be detected normally in the hydrogen spectrum; this line was shown to be produced by an isotope of hydrogen of mass 2, which had been concentrated in the manner described (H. C. Urey, F. G. Brickwedde and G. M. Murphy, 1931). It can be shown theoretically that the frequencies of corresponding lines in the atomic spectra of isotopes should not be identical, but it is only when the ratio of their masses is large, as it is for the isotopes of hydrogen, that the frequency separation is sufficient to make the method useful for the detection of isotopes. The Rydberg constant involved in the equation giving the frequencies of lines in the spectra of atoms (p. 48) contains the term $M/(m + M)$, where m is the mass of an electron and M that of the nucleus of the atom. It follows, therefore, that for elements with nuclear masses M_1 and M_2 , the corresponding Rydberg constants R_1 and R_2 will be related by the expression

$$\frac{R_1}{R_2} = \frac{M_1}{M_2} \cdot \frac{m + M_2}{m + M_1} \quad (24)$$

If the ratio of M_1/M_2 is close to unity, the difference in the Rydberg constants will be extremely small and almost undetectable; for the two isotopes of hydrogen, however, whose masses are approximately 1 and 2, there is an appreciable difference. The Rydberg constant for hydrogen of mass 1 is $109,677.76 \text{ cm.}^{-1}$ (p. 37), and the value for the isotope of mass 2 has been calculated to be $109,707.62 \text{ cm.}^{-1}$. There will thus be a small, but measurable, separation of corresponding lines in the spectrum. The first four lines in the Balmer spectrum and six in the Lyman series have been examined and the frequency differences for the two hydrogen isotopes found to be in agreement with expectation.

The presence of an isotopic atom also produces a shift in the positions of certain lines in the band spectra of molecules. It will be seen in Chapter VIII that the frequency $\bar{\nu}$ of the origin of the vibrational band of a diatomic molecule is given in its simplest form by

$$\bar{\nu} = \nu\omega_e, \quad (25)$$

where ω_e , in cm^{-1} units, is the equilibrium vibration frequency of the molecule, that is, for small amplitudes, and ν is an integer. If the vibration of the two atoms is simple harmonic in nature, then the value of ω_e is related to the restoring force f acting between the atoms by the equation

$$2\pi\omega_e = (f/\mu)^{\frac{1}{2}}, \quad (26)$$

where μ is the reduced mass (cf. p. 47) defined by

$$\frac{1}{\mu} = \frac{1}{m_a} + \frac{1}{m_b}, \quad (27)$$

m_a and m_b being the masses of the two atoms in the molecule. If two isotopic forms of a given molecule are considered the reduced masses will be different e.g., μ_1 and μ_2 , and since the restoring force is probably unchanged, it follows from (25) and (26) that

$$\frac{(\omega_e)_1}{(\omega_e)_2} = \frac{\bar{\nu}_1}{\bar{\nu}_2} = \left(\frac{\mu_2}{\mu_1}\right)^{\frac{1}{2}} \quad (28)$$

$$\therefore \frac{\bar{\nu}_1 - \bar{\nu}_2}{\bar{\nu}_1} = 1 - \left(\frac{\mu_1}{\mu_2}\right)^{\frac{1}{2}}. \quad (29)$$

The frequency shift $\bar{\nu}_1 - \bar{\nu}_2$ of the center of the band thus depends on the reduced masses of the two isotopic molecules, and consequently there may be an appreciable separation which will apply to all the lines in the given band. An isotopic shift similarly occurs in the vibrational bands of the electronic spectrum of a molecule, although the calculation of its extent is here somewhat more complicated. The presence of faint lines in the electronic band spectrum of molecular oxygen led to the discovery of isotopes of masses 17 and 18 (F. W. GIAUQUE and H. L. JOHNSTON, 1929); similarly the isotope of carbon, mass 13, and of nitrogen, mass 15, were first detected by the faint doubling of lines in the spectra of the C_2 molecule and of nitric oxide, respectively. The existence of these isotopes has been confirmed by the mass spectrograph.¹⁴

Results.—The methods described above have been applied to the study of nearly all the known elements; the *stable* isotopes discovered so far are given in Table 22, which does not include the radioactive elements. There are also many unstable isotopes to which reference will be made subsequently.

It can be seen from Table 22 that the majority of elements exist in two or more stable isotopic forms, tin having as many as ten isotopes. In fact, only about twenty elements are simple, and it is of interest to note that all of these, with one possible exception viz., beryllium, have odd atomic numbers. Apart from potassium, elements of odd atomic number do not have more than two isotopes, and these have odd atomic masses, so that the nuclei always contain an even number of neutrons. The only isotope in Table 22 which does not conform to this rule is the potassium isotope of mass 40, but this is not strictly stable for it has a

TABLE 22. THE ISOTOPES OF NONRADIOACTIVE ELEMENTS

Element	At. No.	Isotopes	Element	At. No.	Isotopes
H	1	1, 2	Ma	43	
He	2	3, 4	Ru	44	96, (98), 99, 100, 101, 102, 104
Li	3	6, 7	Rh	45	101, 103
Be	4	9, 10	Pd	46	102, 104, 105, 106, 108, 110
B	5	10, 11	Ag	47	107, 109
C	6	12, 13	Cd	48	106, 108, 110, 111, 112, 113, 114, 116
N	7	14, 15	In	49	113, 115
O	8	16, 17, 18	Sn	50	112, 114, 115, 116, 117, 118, 119, 120, 122, 124
F	9	19	Sb	51	121, 123
Ne	10	20, 21, 22	Te	52	120, 122, 123, 124, 125, 126, 128, 130
Na	11	23	I	53	127
Mg	12	24, 25, 26	Xe	54	124, 126, 128, 129, 130, 131, 132, 134, 136
Al	13	27	Cs	55	133
Si	14	28, 29, 30	Ba	56	130, 132, 134, 135, 136, 137, 138
P	15	31	La	57	139
S	16	32, 33, 34, 36	Ce	58	136, 138, 140, 142
Cl	17	35, 37	Pr	59	141
A	18	36, 38, 40	Nd	60	142, 143, 144, 145, 146, 148, 150
K	19	39, 40, 41	Il	61	?
Ca	20	40, 42, 43, 46, 48	Sm	62	144, 147, 148, 149, 150, 152, 154
Sc	21	45	Eu	63	151, 153
Ti	22	46, 47, 48, 49, 50	Gd	64	152, 154, 155, 156, 157, 158, 160
V	23	51	Tb	65	159
Cr	24	50, 52, 53, 54	Dy	66	158, 160, 161, 162, 163, 164
Mn	25	55	Ho	67	165
Fe	26	54, 56, 57, 58	Er	68	162, 164, 166, 167, 168, 170
Co	27	57, 59	Tm	69	169
Ni	28	58, 60, 61, 62, 64	Yb	70	168, 170, 171, 172, 1/3, 174, 176
Cu	29	63, 65	Lu	71	175, 176
Zn	30	64, 66, 67, 68, 70	Hf	72	(172), 174, 176, 177, 178, 179, 180
Ga	31	69, 71	Ta	73	181
Ge	32	70, 72, 73, 74, 76	W	74	180, 182, 183, 184, 186
As	33	75	Re	75	185, 187
Se	34	74, 76, 77, 78, 80, 82	Os	76	184, 186, 187, 188, 189, 190, 192
Br	35	79, 81	Ir	77	191, 193
Kr	36	78, 80, 82, 83, 84, 86	Pt	78	192, 194, 195, 196, 198
Rb	37	85, 87	Au	79	197
Sr	38	81, 86, 87, 88	Hg	80	196, 198, 199, 200, 201, 202, 204
Y	39	89	Tl	81	203, 205
Zr	40	90, 91, 92, 94, 96	Pb	82	204, 206, 207, 208
Nb	41	93	Bi	83	209
Mo	42	92, 94, 95, 96, 97, 98, 100			

β -ray activity. Among the elements of even atomic number the isotopes of even atomic weight are by far the most common, although several with odd atomic weights, and hence an odd number of neutrons, do exist. It appears, therefore, that the great majority of stable atomic nuclei contain an even number of neutrons, and that an odd number can be tolerated only if the number of protons is even. The type of nuclear structure most preferred is that consisting of even numbers of both protons and neutrons.

Atomic Weights.—In addition to the detection of isotopes, the modern forms of the mass spectrograph have been used to determine their relative masses with an accuracy much greater than is possible in the ordinary determination of atomic weights, especially for the lighter elements. Some of the results obtained in this manner, which are probably correct

TABLE 23. ATOMIC WEIGHTS OF THE LIGHTER ISOTOPES

Hydrogen	{ 1.00813 2.01473	Carbon	{ 12.00398 13.00761
Helium	4.00389	Nitrogen	{ 14.00750 15.00489
Lithium	7.01818	Oxygen	{ 16.00000 18.00369
Beryllium	9.01516	Fluorine	19.00452
Boron	{ 10.01631 11.01292		

to 1 part in 10,000, are quoted in Table 23; the atomic weight of the most abundant isotope of oxygen is taken to be 16.00000.

It is important to emphasize that the standard for atomic weights determined by the mass spectrograph differs from the usual chemical standard; the former is based on a value of 16.0000 for the main isotope of oxygen, whereas the latter assumes this figure to apply to ordinary oxygen, which contains isotopes of masses 17 and 18 to the extent of 0.04 per cent and 0.20 per cent respectively. The mean atomic weight of normal oxygen would consequently be 16.0044 on the mass spectrograph scale; in order to convert the latter values to the usual chemical atomic weight standard it is necessary, therefore, to divide by 16.0044/16, i.e., by 1.00027.

It will be noted that the atomic weight figures in Table 22 are quoted as whole numbers; these are called the **mass numbers** of the isotopes, and are the integers nearest to the actual atomic weights. Although the latter are not exactly integral, as may be seen from Table 23, the differences from whole numbers throughout the whole series of isotopes are so small as to be very remarkable. The largest discrepancy observed is only about 0.1 for the most abundant isotope of tin, which has an atomic weight of 119.91 as determined by the mass spectrograph.

The existence of fractional atomic weights, a fact that had long defied a simple interpretation, is readily explained as being due to mixtures in various proportions of isotopes which themselves have approximately integral atomic weights. By determining the intensities of the lines obtained in the mass spectrum it is possible to estimate the relative amounts of the isotopes normally present in a given element; as the atomic weights of the isotopes are known with some accuracy, the mean atomic weight can be calculated and compared with the generally accepted chemical value. The calculations have been carried out in a large number of cases and excellent agreement found between the two sets of results. For example, boron has two isotopes of atomic weights 10.0163 and 11.0129, and the amounts present in the normal element are 20.0 and 80.0 per cent respectively; the mean atomic weight is, therefore, readily found to be 10.814, and if this figure is divided by 1.00027, to reduce it to the chemical scale, the result is 10.81. The accepted atomic weight determined by chemical methods is 10.82. It is of interest to record, however, that when the isotopic analysis of boron was first carried out, the "physical" atomic weight appeared to be 10.75 ± 0.07 , whereas the "chemical" value was given as 10.90. This discrepancy stimulated a redetermination of the atomic weight of the element by chemical methods with the result given above. With the element antimony also, the mass spectrographic data showed the chemical atomic weight to be inaccurate; it was found to possess two isotopes of mass approximately 121 and 123, present in almost equal amounts, so that the mean atomic weight should be about 122. The recognized value until recently was 120.2, but further investigation showed this to be in error, and the international atomic weight of anti-

mony is now given as 121.76, in excellent agreement with the mass spectroscopic figure of 121.78. The atomic weights of beryllium, krypton, xenon, osmium, and even of hydrogen, have been revised in recent years because of discrepancies between the physical and chemical data.

The discovery of isotopes has simplified the difficulty which has long existed concerning the apparent inversion of certain pairs of elements in the periodic classification. Although the mean atomic weight of argon is greater than that of potassium, the former has two isotopic forms (36 and 38) with smaller masses than the lightest isotope of the latter (39). Further, four of the isotopes of nickel are heavier than those of cobalt, whereas those of tellurium are mostly lighter than iodine, although the mean atomic weights are in the reverse order. It is evident, however, since isotopes are so common, that the continuous increase of the mean atomic weight with increasing atomic number is to be regarded as a fortunate chance, to a great extent, and it is remarkable that the so-called discrepancies are not more common.

Distribution of Isotopes in Nature.—Since actual atomic weights are dependent on the proportions of the isotopes present, it is conceivable that the same element obtained from different sources would have varying mean atomic weights: this is the case, as already seen, with lead of radioactive origin. A similar divergence may exist with calcium, the atomic weight of which in certain ancient rocks appears to be higher than the usual value (J. Kendall, 1933). With chlorine, nickel, silicon and other elements, obtained from widely different sources, however, including specimens of meteoric origin, the atomic weights have been found to be constant within the limits of experimental error.* It would appear, therefore, that in the evolution of the elements the various isotopes, unless arising from radioactive disintegration, have always been produced in the same proportions and these have remained constant. The constancy of the atomic weights of elements from terrestrial sources is not unexpected, since the earth was originally in a molten state in which all the isotopes would be uniformly mixed, but the fact that the same values are obtained with specimens of meteoric origin implies the same definite evolutionary process in both cases. It is possible that the view expressed above may have to be modified because of the discovery that different specimens of lead, not associated with radioactive material, have different isotopic constitutions, although their mean atomic weights are almost identical (A. O. Nier, 1938). The data for two extreme cases are as follows; the lead ore in each case was galena.

Geological Age	Isotope			
	204	206	207	208
Pre-Cambrian ($1,300 \times 10^6$ yr.)	1.48	23.59	22.64	52.29 per cent
Late Carboniferous (230×10^6 yr.)	1.26	27.31	20.00	51.43

In spite of the variations in the relative amounts of the different isotopes the mean atomic weights, calculated from mass spectroscopic data and determined chemically, are almost identical. It will clearly be necessary to make precise isotopic analyses of other elements from different sources before any conclusions can be drawn concerning the evolution of isotopes.

* Hydrogen and oxygen are exceptional in this respect (see p. 158).

The Packing Fraction.—It had been suggested by W. Prout (1815) that the atoms of all elements are ultimately made up of hydrogen atoms; this would require atomic weights to approximate to whole numbers, but since accurate determinations during the nineteenth century and later indicated that this was not the case, the theory was assumed to be unsound. The discovery that all isotopes have masses which are very nearly integral has led to a general acceptance of a modified form of Prout's hypothesis, namely, that atomic nuclei are made up of hydrogen nuclei (protons) and neutrons having almost the same mass. If this view is correct, then atomic weights of isotopes should be approximately, but not exactly, whole numbers; as the masses of both the proton and neutron are greater than unity on the ordinary atomic weight scale, the isotopic masses should be somewhat in excess of integral values. In actual fact the large majority of values, except for the lightest isotopes, are less than whole numbers; the explanation is, presumably, that in the formation of the heavier nuclei from protons and neutrons there is a decrease of mass due to a loss of energy (cf. p. 31). F. W. Aston (1927) has expressed the differences from integers in the form of the **packing fraction** for each isotope, defined in the following manner:

$$\text{Packing fraction} = \frac{\text{Isotopic atomic weight} - \text{Mass number}}{\text{Mass number}} \times 10^4.$$

A negative packing fraction means that the isotopic weight is less than integral, and consequently mass has been transformed into energy in the formation of that particular nucleus. Such nuclei should be particularly stable, whereas a positive packing fraction would imply a tendency to instability. The latter part of this generalization is not strictly correct, especially for elements of small atomic weight, because the masses of the protons and neutrons, of which the nuclei are composed, are somewhat

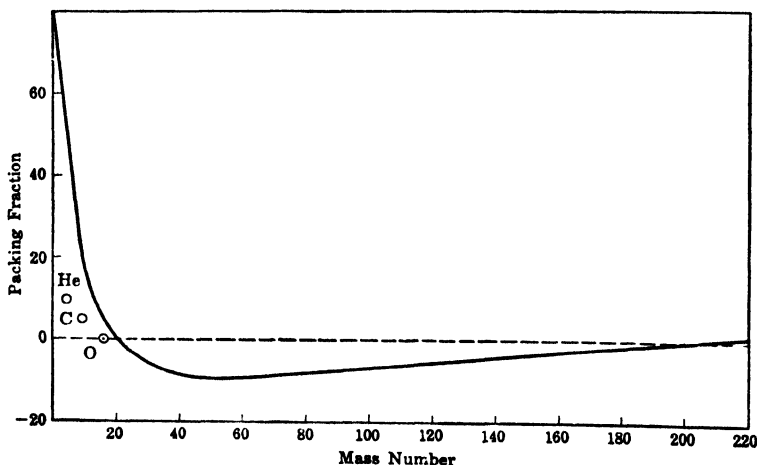


FIG. 33. Packing fractions (Aston)

greater than unity. When plotted against the corresponding mass numbers the packing fractions nearly all fall on or very close to the curve shown in Fig. 33; only helium, carbon (12) and oxygen (16) show any departure from the general behavior. The relatively low packing fractions for these elements indicates exceptional stability; this has been confirmed in another way (p. 161). The lowest packing fractions are observed for the transition elements between chromium and zinc, but with increasing atomic number the fraction increases and for mass numbers exceeding 200 it becomes positive, indicating growing instability. This agrees with the occurrence of spontaneous disintegration, i.e., radioactivity, among such elements; the nuclei are then relatively unstable and so break up until a lighter and more stable nucleus is formed.

A careful reinvestigation of the isotopic weights of the elements of atomic number from 10 to 31 has shown an alternation in the values of the packing fractions when plotted against the respective atomic numbers. The packing fraction of an element with an odd atomic number is greater than the average for the two adjacent elements of even atomic number. Such a result implies that elements with even atomic number, whose nuclei are frequently composed entirely of proton-neutron pairs, are in general more stable than those having odd atomic numbers (H. E. Duckworth, 1942).¹⁵

Isobares.—It may be noted from Table 22 that there are several instances in which different elements, e.g., argon and calcium, have isotopes of the same atomic weight; such isotopic forms have been called **isobares** (Greek: *same weight*) by A. W. Stewart (1918). Isobares, are, of course, of frequent occurrence among the radioelements, since the product of β -ray activity is always an isobare of its parent; the atomic numbers, and hence the chemical properties, differ although the mass number is unchanged. The existence of isobares emphasizes further the secondary nature of the atomic weight, for it is not this but the atomic number that is of fundamental importance in determining chemical characteristics. It will be seen later (p. 171) that there are a number of instances of isobaric isotopes, that is, nuclei with the same atomic mass and atomic number, but differing in radioactive properties. This phenomenon is referred to as "nuclear isomerism."

The Separation of Isotopes: Diffusion Methods.—The preliminary work of Aston, mentioned on p. 142, had shown the possibility of the partial separation of isotopes by diffusion through pipes of baked clay, and the results obtained later by means of the mass spectrograph stimulated further attempts at separating elements into fractions of different atomic weights. Any doubts that existed concerning the interpretation of the observations on positive rays were soon dispelled by the proof that certain, apparently homogeneous, elements could be separated into fractions with different atomic weights.

The rate of diffusion of a gas or vapor is inversely proportional to the square root of its molecular weight (p. 254); the molecules of a lighter isotopic form will thus diffuse through a porous membrane, e.g., pipe-clay tubes, more rapidly than a heavier form and partial separation should be possible. The greater the ratio of the masses of the two isotopic molecules the more efficient

will be the separation, and consequently the method should be more effective with the lighter elements.* The principle was first used by Aston in his attempt to separate the isotopes of neon, and it was also employed by W. D. Harkins (1921, 1926) who worked with hydrogen chloride and obtained two extreme fractions, in different experiments, in which the atomic weight of the chlorine was 35.418 and 35.515, respectively, compared with the initial value 35.457. The former of these evidently contained relatively more of the lighter isotope, mass number 35, and the latter more of the heavier isotope of chlorine (mass number 37) than is present in ordinary chlorine, or in hydrogen chloride. The diffusion method, in the hands of G. Hertz (1932), has proved a very efficient means of separating isotopes. By using twenty-four diffusion pumps in series he was able to obtain specimens of neon containing about 1 per cent and 70 per cent respectively of the neon isotope of mass number 22, whereas the original neon contained about 10 per cent. A more spectacular achievement (1933), however, involving the use of forty-eight diffusion units and mercury vapor as the carrier gas, was the separation of hydrogen of mass number 2 in such a pure form that it did not show the spectrum of the lighter isotope, although at the commencement the gas contained 99.9 per cent of the latter. By means of the diffusion method appreciable separations of the isotopes of carbon, nitrogen, oxygen, neon and argon have also been achieved. The chief disadvantage of the procedure is the small quantity of material which can be handled, on account of the low pressure in the diffusion system; this difficulty has been overcome, to some extent, by P. Capron and M. de Hemptinne (1939) who succeeded in obtaining a concentration of 50 per cent of ^{13}C in methane. These workers introduced a reservoir of methane, in the form of the gas adsorbed on silica gel cooled in liquid air, at the "light" end of the pump circuit.

In spite of the high atomic weight of uranium, the diffusion method has been successfully applied to the partial separation of the isotope of mass number 235, present to the extent of 0.7 per cent, from the predominant isotope of mass number 238. By using uranium hexafluoride (UF_6) as the diffusing gas, in conjunction with special porous barriers as diffusion membranes, and employing some thousands of stages, an enrichment of the lighter isotope of the order of ten-fold has been obtained. This remarkable process has been employed on the large scale to obtain the uranium 235 isotope for the "atomic bomb" (p. 174).

Thermal Diffusion Method.—The procedure described by K. Clusius and G. Dickel (1938) has proved to be of particular interest for the separation of gaseous isotopic molecules; it makes use of a combination of thermal diffusion and of thermal siphoning. The apparatus consists of a vertical tube, 1 to 36 meters in length, along the axis of which is a wire or thin tube heated electrically to about 500° or more; the gas under treatment is placed in the annular space and is thus situated between a hot and a cold surface. As a consequence of thermal diffusion the concentration of heavier molecules is greater in the cold than the hot regions. Further, the thermal siphoning effect causes the gas to rise at the hot wall, to flow down the cold wall after it reaches the top of the column, and then to rise up the hot wall again from the bottom. This continuous streaming of the gas, together with the influence of diffusion, results in a

* The ratio of the molecular weights of the isotopes of hydrogen, mass numbers 2 and 1 respectively, is 2.0, whereas for the isotopes of chlorine, mass numbers 37 and 35, it is only 1.057.

concentration of the heavier component of the gas at the bottom of the tube, while the lighter constituent is found in excess at the top. The method has been employed for the rapid and effective separation of the isotopes of chlorine, neon, carbon and argon.¹⁶

The thermal diffusion principle for the separation of molecules of different masses is applicable to pure liquids and also to solutions. It has been used for the partial enrichment of the 235 isotope of uranium in liquid uranium hexafluoride in one method developed for the production of this isotope in large quantities.

Evaporation and Distillation Methods.—Another successful procedure for separating isotopes depends on the fact that the rates at which otherwise identical atoms or molecules evaporate from a given surface is inversely proportional to the square roots of their masses. If the evaporating molecules are immediately condensed and prevented from returning to the surface from which they originate, the resulting condensate will contain excess of the lighter isotopic forms. J. N. Brønsted and G. Hevesy (1920, 1922) allowed mercury to evaporate at 40° to 60° c., in a highly evacuated space, and condensed the vapor by freezing it on a surface cooled in liquid air placed 1 to 2 cm. above the evaporating mercury. After some time the latter was removed, the frozen mercury allowed to melt, made to evaporate at about 60° c., and the vapor again condensed. This process of fractional evaporation was repeated several times, and eventually fractions having densities of 0.99974 and 1.00023, referred to that of the original mercury as unity, were obtained. The same method was used by O. Hönigschmid and L. Birkenbach (1923) who determined the atomic weights of the lightest and heaviest mercury fractions; they were 200.564 ± 0.006 and 200.632 ± 0.006 , compared with the normal value of 200.61 ± 0.007 . The general nature of the results has been confirmed by W. D. Harkins and his collaborators; they used a special apparatus (1923) made of steel enabling large quantities of mercury to be fractionated rapidly, and obtained (1928) two extreme fractions, each of about 100 g., with an atomic weight difference of nearly 0.19 unit.

The difference in the rates of evaporation was also used by J. N. Brønsted and G. Hevesy (1921) to bring about a fractional separation of the isotopes of chlorine: an aqueous solution of hydrochloric acid was allowed to evaporate at -50° c., and the vapor condensed on a surface cooled in liquid air. By repeated fractionation specimens were obtained in which the average atomic weights of the chlorine differed by at least 0.02 unit. Successful enrichment of the isotopes of zinc (A. C. G. Egerton, 1927) and of potassium (G. Hevesy, 1927) has been achieved by the use of the same principle. In the latter instance the heavier fraction was found to show a stronger β -ray activity than the lighter; it was thought at the time that the radioactivity was to be ascribed to the isotope of mass number 41, but it is now believed to be due to the 40 isotope. Both of these will, of course, concentrate in the heavier fraction.

Although a lighter isotopic molecule leaves a surface more readily than the heavier form, it will have a higher velocity and so will also return more easily, so that the equilibrium vapor pressure (p. 443) may change very little. From theoretical considerations, however, it can be deduced that the lighter isotopic form will generally have a higher vapor pressure, so that a partial separation by normal distillation should be possible. The efficiency of the method will depend on the relative vapor pressures of the isotopic molecules; this should be greatest at low temperatures and will, in any case, only be appreciable for

elements of low atomic weight. The first attempts by F. W. Aston (1913) to separate the isotopes of neon by fractional distillation at liquid air temperatures were unsuccessful, but W. H. Keesom and H. van Dijk (1932) working with a very efficient rectifying column, at temperatures close to the triple point (p. 465), obtained extreme fractions with atomic weights of 20.14 and 20.23, as shown by their densities, compared with 20.18 for ordinary neon. It has already been indicated (p. 147) that a partial enrichment of the heavier isotope of hydrogen was achieved by allowing 4 liters of the liquid to evaporate slowly until only 1 cc. remained, but better results were obtained by W. H. Keesom (1933) with the rectification apparatus employed for the separation of the isotopes of neon. By the fractional distillation of 40 liters of ordinary liquid hydrogen, containing 0.02 per cent of the isotope of mass number 2, he obtained 2 liters of gas containing 1.5 per cent of this form. An appreciable concentration of the heavier isotopes of oxygen has also been achieved by the distillation of water, and partial enrichment of the heavier isotope of nitrogen has been observed in the distillation of liquid ammonia.¹⁷

Positive Ray Method.—The separation of isotopes by adapting the principle of positive ray analysis was first attempted by M. Morand (1926) but without success. In 1934, however, M. L. E. Oliphant, E. S. Shire and B. M. Crowther passed a beam of positively charged lithium ions through crossed electric and magnetic fields, as in the simple mass spectrograph, and collected the two isotopes of lithium, mass numbers 6 and 7, on two small targets placed slightly apart in the calculated positions. Quantities of about 5×10^{-8} g. of each isotope were thus obtained in a pure state. The apparatus was improved by W. R. Smythe (1934 *et seq.*) so as to give more powerful ion beams, and by its means the three isotopes of potassium, mass numbers 39, 40 and 41, were separated, and the middle one shown to possess the β -ray activity of this element. The positive ray method involves many practical difficulties, but it has the advantage of giving a complete separation; it can be applied to heavy as well as light elements. Under the name of "electromagnetic method" the procedure has been employed in the large-scale separation of the isotopes of uranium for use in the "atomic bomb." By utilizing a compound already partially enriched in the 235 isotope by the thermal diffusion method, the efficiency of the separation has been greatly increased.¹⁸

Electrolytic Method.—It was suggested by J. Kendall and E. D. Crittenden (1923) that isotopes might be separated by electrolysis, but there was no evidence as to the value of this suggestion until E. W. Washburn and H. C. Urey (1932) reported that water obtained from an industrial electrolytic cell, which had been in use for some years for the production of oxygen and hydrogen, had a higher density than ordinary water. It appeared, therefore, that on electrolysis of an aqueous solution the lighter isotope of hydrogen is evolved preferentially, leaving an excess of the heavier form in the residual water. This was confirmed by G. N. Lewis and R. T. Macdonald (1933); these workers started with 20 liters of water containing alkali from an old electrolytic cell, and by continued electrolysis obtained a residue of 0.5 ml. of water with a density of 1.073, in which about 70 per cent of the hydrogen was in the form of the heavier isotope. The electrolytic method has been used on a commercial scale to prepare almost pure "heavy water," that is, water in which nearly the whole of the hydrogen has a mass number of 2, as compared with 0.02 per cent in ordinary water. A 0.5 N solution of sodium hydroxide is electrolyzed using a large cur-

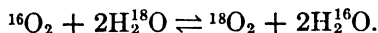
rent generally with nickel electrodes, until all but about one-tenth of the water has been decomposed into hydrogen and oxygen. By burning the gases evolved in the early stages it is possible to obtain water which is almost entirely free from the heavier isotope of hydrogen. The residue in the cell has now become very concentrated in alkali, and so nine-tenths of the liquid is treated with carbon dioxide gas to produce carbonate, and then distilled to dryness. The distillate is added to the remaining one-tenth of electrolyte, thus restoring, approximately, its initial concentration of alkali. The process of electrolysis and removal of alkali is then continued in as many stages as are necessary to leave a residue with the desired concentration of the heavy isotope. During the later stages of the process the hydrogen gas evolved is relatively rich in this isotope, and so it is generally burned and the water condensed and saved for further electrolysis. An indication of the rate of concentration of the heavy isotope can be obtained from the results in Table 24 (H. S. Taylor, H. Eyring

TABLE 24. CONCENTRATION OF HEAVY ISOTOPE OF HYDROGEN BY ELECTROLYSIS

Stage	Volume of Electrolyte	Density	Heavy Isotope
I	2311 liters	0.998	0.03 per cent
II	341 "	0.999	0.5 "
III	52 "	1.001	2.5 "
IV	10 "	1.007	8.0 "
V	2 "	1.031	30.0 "
VI	420 ml.	1.098	93.0 "
VII	82 "	1.104	99.0 "

and A. A. Frost, 1933); the initial electrolyte was from a commercial cell and contained 0.03 per cent of this isotope. As a general rule, to obtain water containing 99 per cent of the heavier isotope it is necessary to electrolyze ordinary water until it is reduced to about one hundred-thousandth of its original volume. It may be mentioned that during electrolysis of water there is also a partial concentration in the residue of the heavier isotopes of oxygen.¹⁹

Chemical Methods.—When isotopes were first discovered, and for many years after, it was generally accepted that they possessed identical chemical properties; this view has undergone modification in recent times. It will be seen in Chapter XIII that isotopic molecules should react at different rates, and in a reversible reaction involving an exchange between two isotopic forms of a given element, the equilibrium constant departs from the simple value to be otherwise expected. Consider, for example, the exchange reaction involving oxygen gas and water, thus *



If the two isotopic forms of oxygen were equally reactive the equilibrium constant of the reaction should be unity, whereas it is actually 1.012 at 25° c.; this result implies that the reaction of the $^{16}\text{O}_2$ molecules with the H_2^{18}O form of water is more rapid than the reverse process.

The differences in the reactivities of isotopic molecules are mainly to be attributed to the differences in the **zero-point energy**, sometimes called

* In order to distinguish between the different isotopes of an element, it is the practice to insert the mass number as a superscript to the symbol.

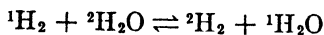
the **residual energy**; this is the energy which, according to quantum mechanics, a molecule must have even at the absolute zero of temperature. For a diatomic molecule, the two atoms of which are supposed to vibrate harmonically with regard to one another, the zero-point energy E_0 per mole is given by (cf. p. 576)

$$E_0 = \frac{1}{2}Nhc\omega_e, \quad (30)$$

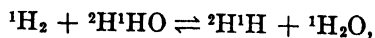
where N is the Avogadro number, h is Planck's constant, c the velocity of light, and ω_e is the equilibrium vibration frequency in cm^{-1} . It was shown on p. 148 that ω_e is related to the reduced mass of the molecule, and hence is dependent on the nature of the isotope present; the magnitude of the zero-point energy thus varies in an analogous manner. The equilibrium constant of an exchange reaction can be shown theoretically, by the methods of statistical mechanics (p. 879), to be determined partly by the molecular weights and moments of inertia of the reacting molecules, but mainly by the difference in the zero-point energies of the substances on the right and left hand side of the chemical equation. It can thus be seen, in a general way, that the equilibrium constant of an isotopic exchange reaction, such as that between oxygen and water, is not necessarily unity. The departure from unity is actually greater the lower the temperature.

One of the consequences of the facts indicated above is that the separation of isotopes by chemical reactions is possible. Since the equilibrium constant in the reaction between gaseous oxygen and liquid water is greater than unity, it follows that the proportion of the heavy isotope ^{18}O in the gas phase is greater than in the liquid phase. If gaseous oxygen, therefore, is passed continuously through liquid water, containing a small amount of the ^{18}O isotope, so that equilibrium is always attained, the resulting gas will contain an increased proportion of the latter. As a general rule a catalyst and a relatively high temperature are necessary to permit equilibrium between gas and liquid to be attained, and so the process has not been used for actual enrichment of the heavier oxygen isotope. It is a remarkable fact, however, that atmospheric oxygen gas is relatively richer in the ^{18}O isotope than is the oxygen in ordinary water, and this difference is attributed to the setting up of the equilibrium under discussion; it is probable that in nature factors are operative, e.g., presence of ultraviolet radiation, which enable equilibrium to be reached at ordinary temperatures.

It has been known for some time that the exchange reaction between hydrogen gas and liquid water, viz.,



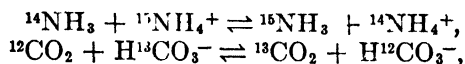
or, in the early stages,



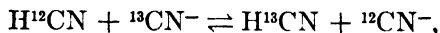
gives at equilibrium a gas in which the ratio of the heavier to the lighter isotope is about three times as great as in the liquid. The attainment of equilibrium has been too slow for the reaction to be of any practical value,

but the development, during the war, of a suitable catalyst (H. S. Taylor) has resulted in the production of heavy water by a large-scale process which is in some ways superior to the electrolytic procedure described above.

When one of the species involved in the isotopic exchange is an acidic or basic ion, it is sometimes possible for relatively rapid reactions to be devised; some examples are the following:



and



involving the exchange between ammonia and ammonium ions, between carbon dioxide and bicarbonate ion,* and between hydrogen cyanide and cyanide ions. These processes have been used for the enrichment of the heavier isotopes of nitrogen and carbon, respectively (H. C. Urey, *et al.*, 1936 *et seq.*).²⁰

Hydrogen Isotopes: Deuterium.—Because the atomic weights of the two isotopes of hydrogen are in the approximate ratio of 2 to 1, which is greatly in excess of that for any other element, the otherwise small differences in physical and chemical properties of isotopes are so greatly accentuated that it has been considered desirable to ascribe different names to them. The hydrogen isotope of mass number two is called **deuterium** (Greek: *the second*), and is given the symbol D; the analogous name **protium** (Greek: *the first*) has been suggested for the lighter isotope, but this has not come into general use. The isotope of mass number unity is represented by H or ^1H and is called hydrogen, and the same name is used for the general description of the element, unless there is any danger of confusion arising. The nucleus of the deuterium atom is called a **deuteron** by analogy with proton. The marked differences in some of the physical properties of the two isotopes and of their oxides, i.e., light and heavy water, respectively, are shown by the data in Table 25.

The relatively large difference in the zero-point energies of molecular hydrogen and deuterium accounts for the difference in the rates at which the two isotopes react (see p. 1104): Hydrogen is, therefore, a particularly good subject for the application of the chemical method of isotopic enrichment; as stated above, the reaction with liquid water has been used for this purpose. Molecular hydrogen exists in three forms, viz., H_2 , D_2 , and HD , the latter containing one atom of each isotope, and the reduced masses (p. 148) are $1/2$, 1 and $2/3$, respectively, assuming the atomic weights to be integral. The zero-point energy of a diatomic molecule is proportional to the equilibrium vibration frequency ω_e by (30), and the latter is inversely proportional to the square root of the reduced mass (26). It follows, therefore, that the zero-point energies of the three

* An enzyme is used to catalyze this reaction.

TABLE 25A. PHYSICAL PROPERTIES OF HYDROGEN ISOTOPES

Property	Hydrogen	Deuterium
Boiling point	20.38° Abs.	23.50° Abs.
Triple point	13.92°	18.58°
Heat of fusion	28 cal./mole	47 cal./mole
Molar volume of solid	26.15 cc.	23.17 cc.

TABLE 25B. PHYSICAL PROPERTIES OF ISOTOPIC FORMS OF WATER

Property	Hydrogen Oxide	Deuterium Oxide
Density at 25°/25°	1.0000	1.1079
Boiling point	100.00°	101.42
Freezing point	0.00°	3.82°
Surface tension	72.75 dynes/cm.	67.8 dynes/cm.

isotopic forms of molecular hydrogen have the following ratios:

$$\begin{aligned} \text{Ratio of Zero-Point Energies} &= \frac{\text{H}_2}{\text{HD}} : \frac{\text{HD}}{\text{D}_2} : \frac{\text{D}_2}{\text{H}_2} \\ &= 1.41 : 1.22 : 1 \end{aligned}$$

The equilibrium vibration frequency of molecular H_2 is known from its band spectrum, and the zero-point energy calculated from it is 6180 cal. per mole; the values for HD and D_2 should thus be 5050 and 4360 cal. per mole, respectively. These results are in agreement with the values obtained by direct measurements on the spectra of HD and D_2 ; they show that the differences between the zero-point energies of the different forms of hydrogen are quite appreciable.

Separation Factor.—The efficiency of a process for the separation of isotopes can be expressed in the form of the **separation factor**; this may be defined as the ratio of the isotopic forms in one phase to that in the other. For example, if $[\text{H}]_l$ and $[\text{D}]_l$ represent the concentrations of the two isotopes of hydrogen in the electrolyte and $[\text{H}]_g$ and $[\text{D}]_g$ the concentrations in the gas obtained by electrolysis, then the electrolytic separation factor s is given by

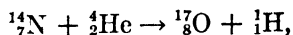
$$s = \frac{[\text{H}]_g}{[\text{D}]_g} \bigg/ \frac{[\text{H}]_l}{[\text{D}]_l}. \quad (31)$$

The value of this factor in the electrolysis of water is approximately 6 for many electrode materials and electrolytes, and represents a comparatively high efficiency of isotopic separation. The same is true of the reaction between hydrogen gas and liquid water referred to earlier; the theoretical separation factor is about 3 at ordinary temperatures, but it decreases as the temperature is raised. For most processes, e.g., the reactions between oxygen and water, or between ammonia and ammonium ions, the separation factor is little more than unity, indicating that the enrichment obtained is small.²¹

ARTIFICIAL DISINTEGRATION OF ATOMS

Disintegration by Alpha-Particles.—The type of nuclear disintegration considered in connection with radioactivity involved the spontaneous breaking down of the atomic nuclei; it has been found possible, also, to bring about various types of nuclear transformations by artificial means. In 1919, E. Rutherford noted that when α -particles of high velocity, e.g., from radium-*B* and -*C*, passed through nitrogen gas, protons of long range were produced. These protons could not have come from hydrogen present as an impurity, for the maximum range of the particles obtained by passage of the α -rays through hydrogen gas was 28 cm., compared with ranges up to 40 cm. for those from nitrogen. In later work by E. Rutherford and J. Chadwick (1921, 1922), and by others, it was found that fast moving protons could be obtained, as a result of α -ray bombardment, from all elements up to calcium, with the exception of helium, carbon and oxygen. The exceptional behavior in this respect of the latter three elements provides confirmation of the view already expressed (p. 153) that the atoms ${}^4\text{He}$, ${}^{12}\text{C}$ and ${}^{16}\text{O}$ are very stable. On account of the very powerful repulsion between an atomic nucleus and the positively charged α -particle, which increases very rapidly with increasing charge on the former (see p. 24), no interaction is to be expected between elements of relatively high atomic number and α -particles obtained from radioactive sources. By the use of helium nuclei specially accelerated by methods described below for protons and deuterons, however, it is possible to eject protons from several of the heavier elements, including titanium, vanadium, chromium and nickel.²²

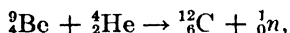
The production of protons, and other particles, as a result of α -ray bombardment has been studied by observation of the tracks obtained in a Wilson cloud-chamber, as well as by the scintillation method, and the conclusion has been reached that as the result of certain collisions between an atomic nucleus and a fast moving α -particle, a proton is ejected from the former, so that an artificial disintegration has been achieved. Actually only a relatively small number of nuclei undergo any change, for there are no more than about twenty fruitful collisions for every million α -particles of 7 cm. range sent into the nitrogen. From a study of the tracks obtained in the Wilson chamber and from considerations based on the conservation of mass and energy (p. 8), it appears that the loss of the proton is not the result merely of an impact between the atomic nucleus and the α -particle, but that the latter enters the nucleus and so forms an unstable system which *immediately* ejects a proton in the effort to attain stability. The disintegration of the nitrogen nucleus may thus be represented by the equation



where each symbol represents the nucleus of the corresponding element, the upper prefix giving, as usual, the mass number and the lower the nuclear charge, i.e., atomic number. It is essential, of course, that the

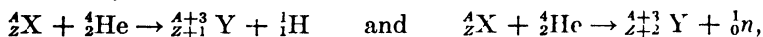
sum of the mass numbers, as well as of the positive charges, should be the same on both sides of the disintegration equation. It will be noted that in the disintegration process a nitrogen atom has been converted into one of oxygen; this consequently represents an instance of the **artificial transmutation** of elements. Many further examples will appear shortly, but the extent of the transformation is so very small that it is rarely possible to identify the product by direct experiment; its nature must generally be inferred in other ways.

It was stated on p. 29 that α -particles are able to bring about a type of disintegration leading to the production of neutrons; this phenomenon is particularly marked with beryllium, although it has been observed with other elements, including several of high atomic number, if specially accelerated helium nuclei are used. The production of neutrons from beryllium may be written



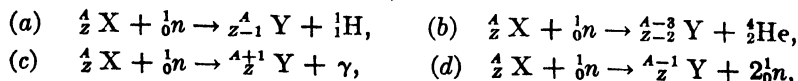
where the symbol 1_0n is used for a neutron.

The two types of disintegration brought about by α -particles may be represented by the general equations

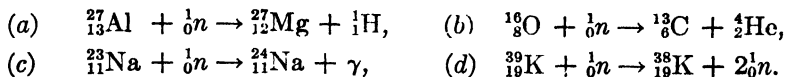


where A and Z are the mass number and atomic number, respectively, of the bombarded element X ; the values for the product Y depend on the nature of the particles produced. The type of disintegration process is often represented in a simple manner by the use of an abbreviation for the projectile particle followed by one for the emitted particle; thus, the two reactions above would be indicated by $(\alpha; p)$ and $(\alpha; n)$, respectively. A specific nuclear reaction would then be represented by preceding this by the symbol of the bombarded element and following it by that of the product; for example, the production of neutrons from beryllium, as mentioned above, would be written ${}^9\text{Be} (\alpha; n) {}^{12}\text{C}$.

Disintegration by Neutrons.—Since neutrons carry no electric charge they are not repelled by atomic nuclei, and so are very effective in producing disruption. In fact, neutrons which have been slowed down by passage through a substance containing a large proportion of hydrogen, e.g., water or paraffin wax, are in many cases more useful in this respect than those of high velocity. The source of neutrons is generally a small tube containing a beryllium salt and radon; the α -particles emitted from the latter, and from its radioactive products, disintegrate the beryllium atoms, as shown above. It will be seen later that very powerful neutron beams can be obtained in other ways, but special apparatus is required. Nearly all the known elements, in one isotopic form or another, interact with neutrons, and four main types of behavior have been observed. These results in the emission of (a) a proton, (b) an α -particle, (c) γ -radiation, and (d) two neutrons; thus,

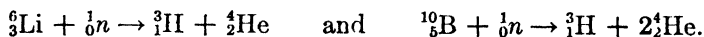


or $(n; p)$, $(n; \alpha)$, $(n; \gamma)$ and $(n; 2n)$, respectively. The processes (a), (b) and (d) usually require fast neutrons and are observed mostly with elements of low atomic weight, but the reaction (c), which almost invariably leads to the formation of a relatively unstable nucleus that breaks up at a definite rate, occurs best with slow neutrons and takes place with the great majority of elements. Actual examples of the four different types of neutron interactions are as follows:



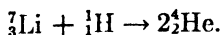
The products of processes (c) and (d) are seen to be isotopes of the bombarded element, although their mass numbers are a unit higher and lower, respectively.²³

A few exceptional results in the bombardment with neutrons may be mentioned; the ^{12}C nucleus is split up into three α -particles, whereas ^6Li and ^{10}B give, in addition to one and two α -particles, respectively, a nucleus of an isotope of hydrogen of mass number 3; thus,



The ^3H isotope, produced in this and other reactions, is of particular importance, as will be seen later (p. 172). In view of the difficulty of proving by direct methods the nature of the products, because they are obtained in such small amounts, it is of interest to note that after bombardment of methyl borate by neutrons for seven weeks, sufficient helium was liberated to be detected spectroscopically (F. Paneth, 1935).

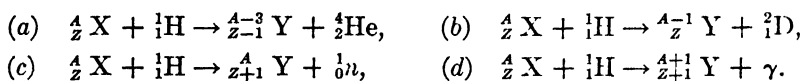
Disintegration by Protons and Deuterons.—An important advance in the study of artificial disintegration was made when it was discovered that atomic nuclei could be broken up by means of protons and deuterons, produced in a discharge tube, which were given a high velocity by passage through a high potential. J. D. Cockroft and E. T. S. Walton (1932) used the "voltage multiplier" principle to obtain these high potentials—condensers were charged in parallel and discharged in series so as to step up the available voltage—and observed the disintegration of lithium by fast protons. A layer of lithium oxide was bombarded by the protons after passing through potentials up to 600,000 volts, and scintillations were observed on a zinc sulfide screen which could not have been due to protons because they were produced by particles of much longer range. From the brightness of the scintillations, and the brightness of the tracks produced in a cloud chamber, it is evident that the lithium nucleus first captures a proton and then breaks up into two helium nuclei which are ejected in opposite directions; the process may be represented by



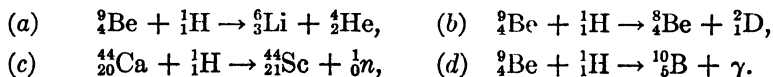
Further progress in the study of disintegration by protons, and also by deuterons which are more effective than protons of the same voltage (G. N. Lewis, E. O. Lawrence and M. S. Livingston, 1933), has depended mainly on

important improvements in the technique of production of strong ion beams of high voltage. Progress in the latter connection has taken place in a number of directions, two of which have attracted particular attention: these are the "magnetic resonance" method of E. O. Lawrence and M. S. Livingston (1933), whose apparatus is generally known as the *cyclotron*, and the electrostatic method of R. J. Van de Graaff (1931, 1936). In the former the apparatus is so arranged that the ions are made to move in a flat spiral of steadily increasing radius, and are given a voltage impulse of about 30,000 to 90,000 volts in each of the several hundred semicircular paths, so that maximum potentials of the order of 100 million volts are possible. The principle of the electrostatic method is virtually the same as that of the familiar Wimshurst machine; the charge is transferred from a high voltage rectifier to a moving belt and so is carried to a large hollow collecting sphere where the potential is built up.

Fast moving protons have been found to bring about at least four types of disintegration, giving (a) α -particles, (b) deuterons, (c) neutrons and (d) γ -rays, respectively; they may be represented as follows: (p ; α), (p ; d), (p ; n) and (p , γ). The interactions are

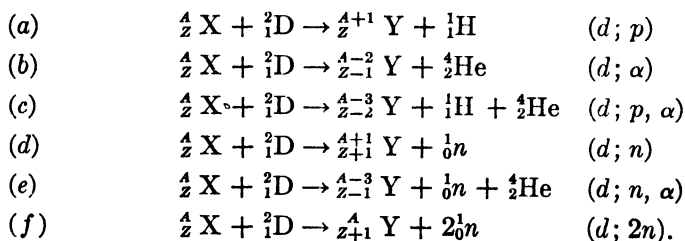


Instances of the four types of disintegration are the following; it is seen that the same isotope may take part in a number of different processes.

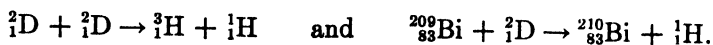


It may be noted that the example quoted under (b) is the only one so far observed in which deuterons are produced as a consequence of proton bombardment. In addition to the cases already recorded, mention should be made of the complete disintegration into three α -particles by the action of a proton on ^{11}B .

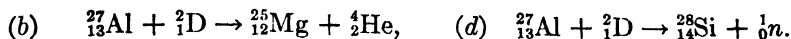
The disintegrations produced by accelerated deuterons fall into six main categories; they are



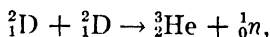
Processes of type (a) are most common and have been detected for light and heavy elements, from deuterium itself to bismuth; thus,



The production of α -particles, reaction (b), generally occurs with light elements, whereas neutron formation, reaction (d), occurs most easily with the heavier elements; examples are



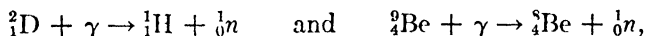
The action of fast deuterons on targets of lithium or beryllium yields neutron beams of great intensity, actually several hundred times as strong as those given by the radon-beryllium method; in this way a powerful means has become available for the investigation of disintegrations by neutrons. Many of the reactions are of considerable interest, as will be evident shortly. Processes (c) and (e), in which two particles are emitted, are rare; platinum is said to give a neutron and an α -particle, and gold a proton and an α -particle. Reaction (f) is also not very common, an example being ${}^{56}_{26}\text{Fe} (d; 2n) {}^{56}_{25}\text{Mn}$. In addition mention may be made of the following disintegrations which do not come within the scope of the six types noted above: ${}^6_3\text{Li}$, ${}^{10}_5\text{B}$ and ${}^{14}_7\text{N}$ give respectively two, three and four α -particles only, whereas ${}^7_3\text{Li}$ and ${}^{11}_5\text{B}$ give a neutron in addition to two and three α -particles, respectively. A specially interesting interaction, alternative to one already mentioned, occurs between two deuterons, viz.,



when the residual nucleus is an isotope of helium, mass number 3, which is present in ordinary helium to the extent of about one part in a million.

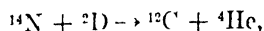
In many of the disintegration reactions involving fast protons or deuterons the excess energy released often appears in the form of γ -radiation.²¹

Disintegration by Gamma-Rays.—A few instances have been recorded of the breakdown of nuclei by the action of γ -radiation of high energy, particularly those obtained from thorium-C'' (2.62×10^6 e.v.), a neutron being ejected. The two best known examples are



the former being of special interest as it leads to the best value for the mass of the neutron, as shown below. A few other elements of low and high atomic weights have been disintegrated by energetic γ -rays with the emission of a neutron.

Conservation of Mass and Energy.—An examination of the mass and energy changes involved in disintegration processes has frequently been used to confirm the nature of the interaction. This subject is of importance as it brings out very clearly the equivalence of mass and energy, as required by the theory of relativity (p. 7). As previously noted, in ordinary chemical reactions the mass equivalent of the energy change is so small as to be negligible, but this is not so for the nuclear reactions under consideration. In the process



the masses of the interacting species, as given in Table 23, are 14.00750 and

2.01473, whereas those of the products are 12.00398 and 4.00389; there is consequently a loss of 0.01436 mass unit.* The energy equivalent of a mass m is mc^2 , where c is the velocity of light, 3×10^{10} cm. per sec.; hence it follows that the mass equivalent of a particle of *relative* mass 1 is given by c^2/N ergs, where N is the Avogadro number, $1/N$ being the *actual* mass of the particle. To convert this into e.v. units (see p. 6) it is necessary to multiply by $300/e$, where e is the electronic charge in e.s. units; hence,

$$\begin{aligned} 1 \text{ mass unit} &\equiv 300 \, c^2/Ne \text{ e.v.} \\ &\equiv 0.93 \times 10^9 \text{ e.v.} \end{aligned}$$

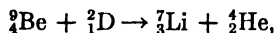
It follows, therefore, that if the disintegration occurs as represented above, and mass and energy are equivalent, the 0.01436 mass unit lost should appear as 13.36×10^6 e.v. of energy. The energy of the bombarded nitrogen nucleus is only kinetic and is negligible in comparison with that of the other particles involved; that of the bombarding deuteron is 0.575×10^6 e.v. from the applied potential, and of the resulting helium nucleus it is 10.41×10^6 e.v. as estimated from its range, after applying various corrections. The recoil energy of the resultant carbon nucleus can be estimated by assuming the conservation of momentum, thus, remembering that the deuteron and the α -particle proceed in the same direction, whereas the carbon nucleus *recoils* in the opposite direction, it follows that

$$m_3 E_3 = m_1 E_1 + m_2 E_2, \quad (32)$$

where m and E are the mass and energy, and the suffixes 1, 2 and 3 refer to the carbon, deuteron and α -particle, respectively. Since all the quantities except E_3 are known, this is readily evaluated as 3.57×10^6 e.v. It follows, therefore, that the excess energy of the disintegration is $(3.57 + 10.41 - 0.575) \times 10^6$, i.e., 13.40×10^6 e.v., in excellent agreement with that estimated from the mass lost in the process.

The disintegration ${}^2_1\text{D} + \gamma \rightarrow {}^1_1\text{H} + {}^1_0\text{n}$ can be used to give accurately the mass of the neutron. The energy of the γ -radiation from thorium-C'' is 2.62×10^6 e.v., and that of the resulting proton and neutron has been estimated to be 0.45×10^6 e.v.; hence, 2.17×10^6 e.v. have disappeared, and have been presumably converted into 0.00233 mass units, in the process. Taking the accurate mass spectroscopic values for the atomic weights of deuterium and hydrogen given in Table 23, the mass of the neutron is $2.01473 + 0.00233 - 1.00813$, i.e., 1.00893 on the *physical* atomic weight scale.

Revision of Atomic Weights.—A remarkable result of the study of mass and energy changes in disintegration processes was the indication that the mass spectroscopic atomic weights in use prior to 1935 were in error; this was pointed out independently by H. Bethe (1935) and by M. L. E. Oliphant, A. E. Kempton and Lord Rutherford (1935). Consider, for example, the process



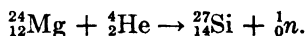
in which the net increase of energy is 7.19×10^6 e.v.; according to the older atomic weights, however, the loss of mass should have been equivalent to 11.4×10^6 e.v. Similar discrepancies were found in other instances, and the

* Although nuclei are involved in the reaction the weights here used are those of the whole atom, but since the number of electrons is the same on each side of the equation the correct result for the mass *difference* is obtained.

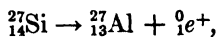
bold suggestion was made that the atomic weights of the lighter isotopes, at least up to ^{14}N , were in error; new observations with the mass spectrograph confirmed this suggestion, and with the newer values the discrepancies have almost entirely disappeared. In the reaction considered the energy equivalent of the mass lost is found to be 7.27×10^6 e.v., in satisfactory agreement with that observed.²⁵

INDUCED RADIOACTIVITY

Unstable Nuclei.—It has been indicated that the products of artificial disintegration are often nuclei which, although not highly unstable, nevertheless break up at a definite rate; this discovery of such isotopes is one of the most remarkable scientific events of recent years. While studying the effects of α -particles on boron, magnesium and aluminum, Mlle. I. Curie and F. Joliot (1934) found that there were two types of disintegration, the main one leading to the emission of protons and the other to the formation of neutrons and positrons, that is, positive electrons (p. 21). On removing the source of α -particles, the neutron and proton emission ceased, but positrons continued to be evolved for some time, the intensity falling off exponentially. When fresh material was exposed to α -particles the activity similarly increased in an exponential manner from zero to the maximum value. It appeared, therefore, that the production of neutrons and positrons, although apparently connected, does not occur simultaneously but must involve two stages which are fundamentally different. The most probable explanation is that the α -particle is first taken up by the bombarded nucleus to form a complex which disintegrates instantaneously with the emission of neutrons; thus, with magnesium,



The product ^{27}Si , called radio-silicon, is not a stable isotope of silicon, but it is not so unstable as to be incapable of existence for any appreciable time. Actually, it disintegrates spontaneously at a definite rate and in an exponential manner exactly analogous to that of an ordinary radioactive element, except that positrons are emitted instead of α - or β -particles; thus,



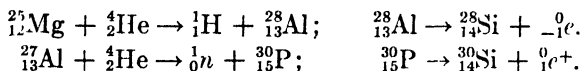
the symbol $^0_1e^+$ being used for the positive electron. The final product is seen to be ^{27}Al , the stable isotope of aluminum (see Table 22).

The results described above show that as a consequence of bombarding certain elements with α -particles it is possible to produce artificial radioelements. Further investigation of the phenomenon of **induced radioactivity**, or **artificial radioactivity** as it is sometimes called, has led to the discovery of some five hundred different unstable isotopic species of the familiar elements, from helium to uranium, with very few exceptions. These have all been obtained by the bombardment of stable elements with α -particles, protons, deuterons, neutrons or γ -rays. In fact, in any of the types of atomic disintegration already considered the product may

be an artificial radioelement. The $(p; d)$ and $(p; \alpha)$ processes, however, usually result in the formation of stable elements. Although the earliest artificial radioelements to be discovered possessed positron activity, many, probably the majority, and especially those of high atomic weight, emit electrons, i.e., β -particles. Frequently the same active isotope can be produced in several different ways; for example, ^{30}P is obtained in the following five disintegrations: $(\alpha; n)$ on aluminum, $(p; n)$ on silicon, $(n; 2n)$ and $(\gamma; n)$ on phosphorus, and $(d; \alpha)$ on sulfur. The products all have the same half-life period, namely, 2.5 min., irrespective of their origin. The half-life periods of the artificial radioelements vary from a fraction of a second to several years, but for most of them the period is of the order of minutes; it is easy to understand, therefore, that these unstable isotopes do not occur naturally.²⁶

Methods of Investigation.—The study of artificial radioactivity falls into two categories, namely, physical and chemical. In the first, the important point is the determination of the rate of decay of the activity, so that the half-life may be calculated. The Geiger-Muller counter (p. 120) is frequently employed on account of its simplicity, but electroscopes are also used for this purpose. The nature of the active isotope can often be determined, or confirmed, by carrying out a chemical separation in the presence of appreciable amounts of the ordinary inactive elements. For example, if aluminum is exposed to α -rays and then dissolved in hydrochloric acid, the evolved hydrogen carries with it the positron activity, probably as phosphine; further, if to the acid solution sodium phosphate and a zirconium salt are added, the precipitate of zirconium phosphate carries all the activity with it. The active element is thus presumably an isotope of phosphorus. Similarly, when carbon is bombarded by deuterons the product is believed to be radio-nitrogen; this has been proved by showing that the active species is separable from every element of the first period of the periodic table except nitrogen. Many other radioelements, e.g., active forms of carbon, oxygen, fluorine, aluminum, chlorine, manganese, arsenic and bromine, have been identified in analogous ways.

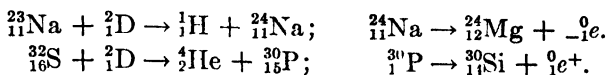
Induced Radioactivity by Alpha-Ray Bombardment.—Both types of α -ray activity, viz., $(\alpha; p)$ and $(\alpha; n)$ give rise to radioactive nuclei; the $(\alpha; p)$ process generally yields stable nuclei, but the few unstable ones produced in this manner usually emit negative particles. The $(\alpha; n)$ interaction apparently leads to active elements emitting positrons only; examples of the two types are



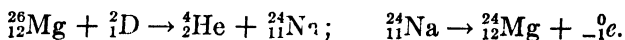
Both ^{28}Si and ^{30}Si are stable isotopes found in ordinary silicon. With a few exceptions (see p. 161), only light elements respond to α -ray bombardment, but titanium, chromium, copper and bromine among others, give radioelements when artificially accelerated helium nuclei are employed.

Induced Radioactivity by Proton and Deuteron Bombardment.—With protons the $(p; n)$ and $(p; \gamma)$ types of disintegration, especially the former,

often lead to the formation of radioactive elements; as already mentioned, the ($p; d$) and ($p; \alpha$) disintegrations usually give stable isotopes. When fast-moving deuterons are used for bombardment purposes radioactive products are common. The disintegrations ($d; p$) and ($d; \alpha$) frequently yield active species, and although both positron and negatron emitters have been obtained, the former process usually gives negative, and the latter positive, particles; examples are the following:

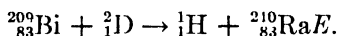


The second type of disintegration may give rise to products having negative electron activity; thus,



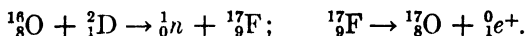
Radio-sodium, ${}^{24}\text{Na}$, which results from the deuteron bombardment of either sodium or magnesium, has a half-life period of 14.8 hrs., and its disintegration is accompanied by both β - and γ -ray activity; it has been obtained in amounts having activities of the same order as natural radioactive materials.

The nature of the disintegration resulting from the impact of deuterons on bismuth has been confirmed in an interesting manner. The product was found to have a β -ray activity with a half-life period of 5.0 days, and this was followed by an α -particle emitter apparently identical with polonium (radium- F). The first unstable element is evidently radium- E , and since bismuth possesses only one stable isotope ${}^{209}\text{Bi}$, the action of deuterons must be



In this manner a natural radioactive element has been produced artificially.*

The ($d; n$) disintegration also gives rise to unstable isotopes; these are always positron emitters; for example,

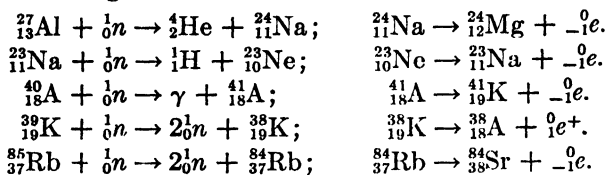


The active element has been shown to be fluorine by chemical methods. Oxygen, in the form of water, was exposed to fast deuterons and then poured into a boiling solution of potassium fluoride; addition of calcium chloride gave a precipitate of calcium fluoride which possessed all the positron activity. The rates of decay of the water after exposure to deuterons and of the calcium fluoride precipitate both indicated a half-life period of 1.16 min., and so were due to the same element, presumably fluorine.

Induced Radioactivity by Neutron Bombardment.—By far the most important results in the field of induced radioactivity have been obtained

* Some radium- F is formed at the same time by the reaction ${}^{209}\text{Bi} (d, n) {}^{210}\text{Ra}F$.

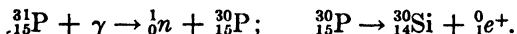
by the use of neutrons as projectiles. It was discovered by E. Fermi and his co-workers (1934, 1935) that the majority of elements gave unstable isotopes when bombarded with neutrons which had been slowed down by passage through water or paraffin wax. This work has been confirmed and extended by many other investigators. The radioelements resulting from the $(n; \alpha)$, $(n; p)$ and $(n; \gamma)$ types of disintegration are always negative electron emitters, whereas the $(n; 2n)$ process leads to both positron and electron activity. The following instances of the various interactions illustrate the general nature of the changes; two examples of $(n; 2n)$ disintegrations are given.



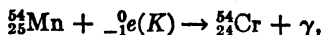
The $(n; \gamma)$ type of activation is somewhat more common than the others.

In view of the ease of production of relatively good supplies of neutrons, by the radon-beryllium method or by the interaction of deuterons, and the frequency with which active elements are obtained with the aid of these particles, neutrons have been widely used for the production of radioelements for a variety of chemical and physical purposes.²⁷

Induced Radioactivity by Action of Gamma-Rays.—It was mentioned on p. 165 that highly energetic γ -rays are able to eject neutrons from a few atomic nuclei; the resulting products may then be radioactive. By the use of the γ -rays of 17×10^6 E.v. energy, accompanying the disintegration of lithium by high velocity protons, it has been found possible to eject neutrons from phosphorus, copper, bromine, silver, and other nuclei, leaving radioactive isotopes; thus,



K-Electron Capture and Internal Conversion.—In addition to the radioactive disintegrations accompanied by the emission of α -particles, β -particles (negative electrons) or positrons, there are two other types which are of somewhat less frequent occurrence. In some processes, when the emission of a positron would be expected, that is, the unstable nucleus has too many protons, there may not be sufficient energy available for this to occur. As an alternative, the nucleus acquires stability by taking up an electron from the *K*-shell; this phenomenon is called *K-electron capture*. The vacancy in the *K*-shell that results is filled by an electron from a higher level, and this will be accompanied by the emission of the *K*-series of characteristic X-rays of the disintegration product (cf. p. 71). The latter will evidently be an element of the same atomic weight but one unit less in atomic number than the parent. An illustration of *K*-electron capture is provided by ${}^{54}\text{Mn}$ produced, for example, by the process ${}^{53}\text{Cr} (d; n) {}^{54}\text{Mn}$; the disintegration is then represented by



where part of the γ -rays are the characteristic X-rays of the ${}^{54}_{24}\text{Cr}$ atom.

The second novel type of radioactive change takes place when part of the energy which might have been expected to be evolved as γ -radiation is utilized to expel an extranuclear electron from the radiating atom. This change is referred to as **internal conversion** of the γ -ray. The electrons emitted have virtually the same energy, and so they are called a "line of electrons," to distinguish them from the β -ray electrons which form a "spectrum" (p. 124). A number of artificial radioelements, e.g., ^{44}Ca , ^{80}Br , ^{133}Ba and ^{180}Ta , exhibit the internal conversion effect. It may be mentioned that some elements decay by simultaneous K -electron capture and internal conversion.

Nuclear Isomerism.—The possibility that two elements might have the same atomic weight as well as the same atomic number, but differ in their radioactive properties, was first realized in connection with the natural radioelements uranium-Z and uranium-X₂ (p. 133). With the development of the study of artificial radioactivity it has become apparent that there are more than twenty pairs of such isobaric isotopes, now known as **isomeric nuclei**; the general phenomenon, which is referred to as **nuclear isomerism**, is probably more common than is realized. There is evidence for the occurrence of different energy levels of an atomic nucleus, and such levels represent different isomeric states. Usually, the transition from an upper, i.e., excited, to the lowest, i.e., ground, state of the nucleus is so rapid that the excited states have virtually no independent existence. If, however, the energy difference between the excited state and the ground state is relatively small and, in addition, there is a considerable difference of angular momentum, i.e., nuclear spin, of the nuclei in the two states, the transition from the upper to the lower state will be "forbidden"; in other words, the transitions will be fairly uncommon, and the excited state will be metastable and have a long half-life period. When this occurs the nuclear state of higher energy can have an appreciable independent existence, and nuclear isomerism will be evident (C. F. von Weizsäcker, 1936). In illustration reference may be made to radio-zinc, ^{69}Zn , obtained by the process $^{68}\text{Zn} (d; p) ^{69}\text{Zn}$; the excited state formed in this manner decays with a half-life of 13.8 hr., emitting γ -radiation, part of which is internally converted. The product is the isomeric nucleus ^{69}Zn , presumably in the ground state, and this gives off electrons, its half-life period being 57 min. A few cases are known in which a stable element has an active isomeric form; for example, ^{83}Kr , produced in various bombardment processes or by the β -ray decay of ^{83}Br , possesses activity, although it has the same atomic weight as ^{83}Kr which is present to the extent of 11.5 per cent in atmospheric krypton. The radioactive isomer is presumably in an excited nuclear energy state, and it decays with a half-life of 113 min. The excess energy, which should appear as γ -radiation, is internally converted, so that a "line" of electrons is emitted.

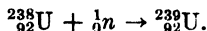
A particularly interesting case of nuclear isomerism is provided by ^{80}Br , where a chemical separation of the excited (half-life, 4.4 hr.) and the normal (half-life, 18 min.) states has been achieved. The γ -ray which should accompany the transition from upper to lower state is internally converted, and the loss of the electron leaves the product virtually in the ionic form. As a result it is especially reactive chemically, and this makes possible a separation from the parent isomer. For example, sodium bromide containing ^{80}Br , prepared by the process $^{81}\text{Br} (n, 2n) ^{80}\text{Br}$, is converted into *tert.*-butyl bromide, which is dissolved in aqueous methanol. Normally the interaction, leading to

the formation of hydrobromic acid, is slow, but with the butyl bromide containing the radioactive ^{80}Br there is a relatively rapid reaction. The ^{80}Br present in the unreacted butyl bromide extracted from the solution is found to decay with a half-life of 4.4 hr., while that in the hydrobromic acid has a half-life of 18 min.²⁸

Radioelements of Special Interest.—As indicated on pp. 163 and 164, an isotope of hydrogen of mass 3, i.e., ^3H , is formed in the neutron bombardment of lithium and boron, and by the interaction of two deuterons. This isotope, which has not been detected in ordinary hydrogen, is radioactive, emitting β -particles, with the exceptionally long half-life period of about 31 years (L. W. Alvarez and R. Cornog, 1939). The ^3H isotope, called **tritium**, symbol T, has also been obtained by the bombardment of nitrogen by neutrons, and of beryllium by deuterons.

By means of artificial radioelements all the gaps in the periodic table, with but one exception, have now been filled. The discovery in nature of the elements of atomic number 43 and 61, called masurium and illinium, respectively, has been reported, but the results were not very decisive. Both these elements, in several radioactive isotopic forms, have now been definitely identified; element 43 results from the $(p; n)$ and $(d; n)$ processes on molybdenum, while 61 is obtained from neodymium by the reactions $(p; n)$, $(d; n)$ and $(\alpha; p)$. The missing element of atomic number 85, mass number 211, a member of the halogen group, has been obtained by the process $(\alpha; 2n)$ on bismuth (^{209}Bi). By adding the product to various compounds of known elements, subjecting the mixtures to chemical reagents, and utilizing the radioactivity (α -ray) to follow the behavior of the element number 85, many of its properties have been determined. In this way it has been shown to behave like a metal, rather than as a halogen. This result is not unexpected in view of the steady increase in electropositive properties in the halogen group with increasing atomic weight. The only element hitherto not definitely identified is that of atomic number 87; it has been suggested as a product of a branched α -ray disintegration of actinium ($^{227}_{89}\text{Ac}$), but this is not certain.²⁹

Fission of Uranium Atoms.—The bombardment of uranium compounds by neutrons was found by E. Fermi (1934, 1935) to give rise to active products emitting electrons; the rate of decay was complex, however, indicating the presence of several radioactive atoms with different half-life periods, and in fact ten such periods were subsequently identified. If the primary disintegration stage is of the $(n; \gamma)$ type, as is probable, the product is an isotope of uranium; thus,

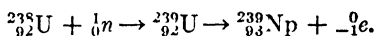


If this new form of uranium gives off a β -particle, the resulting nucleus should have an atomic number of 93, a figure higher than that of any element existing in nature; further, the detection of several different half-life periods would suggest the emission of a number of electrons in stages, thus leading to the formation of a series of so-called "trans-uranic" elements of high atomic number. This interpretation of the observations was at first generally accepted and by studying the separability of these radioactive elements from known elements, with which they were presumably homologous, O. Hahn *et al.* (1937) claimed to have identified a series of elements with atomic numbers from 93 to 97. As a result of the observation by I. Curie and P. Savitch (1938) that one of the half-life decay periods following the bombardment of uranium by neutrons re-

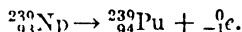
sembled that of radio-lanthanum, O. Hahn and F. Strassmann (1939) re-examined the problem and succeeded in showing that lanthanum and barium could be definitely identified among the products. These conclusions have been confirmed by others, and the detection of various elements, in radioactive forms, whose atomic numbers lie between 35 (bromine) and 57 (lanthanum) has been claimed. It appears, therefore, that after taking up a neutron the resulting uranium nucleus is so unstable that it breaks up immediately into two nuclei of approximately equal size. This process, known as **nuclear fission**, results, to some extent, from the bombardment of the predominant ^{238}U isotope by fast neutrons, and also from the action on the less common isotope ^{235}U of both fast and slow neutrons, the latter being more effective. Because the total mass of the products is less than that of the interacting particles, considerable amounts of energy are liberated during fission, viz., about 200×10^6 e.v. per uranium atom. At the same time, three or four fast neutrons are liberated which might cause the fission of other uranium nuclei, thus producing more energy and more neutrons. There is thus the possibility of the extremely rapid, or explosive, liberation of enormous amounts of energy, as the result of a self-propagating, branching chain of nuclear fissions, initiated by a single (effective) neutron.

Further experiments have shown that the fission of uranium can be brought about by deuterons and α -particles, and also by γ -rays of high energy. The thorium nucleus undergoes a similar fission when bombarded by neutrons, deuterons or γ -rays.

Plutonium: The Atomic Bomb.—In addition to undergoing fission, to some extent, by the action of fast neutrons, it has been found that there is considerable absorption of "resonance" neutrons of certain specific energies, leading to the formation of ^{239}U , as indicated above. The resulting isotope emits a β -particle, the half-life period being 23 min., and the product is an element of mass number 239 and atomic number 93, which has been named **neptunium** and given the symbol Np. The formation of this true "trans-uranic" element may be represented by



Neptunium is itself also a β -particle emitter, with a half-life of 2.3 days, and the product, mass number 239, atomic number 94, is known as **plutonium**, symbol Pu; thus,



Plutonium 239 is a relatively stable element; it emits an α -particle, and its half-life period is 24,000 years. Both neptunium and plutonium have chemical properties resembling those of uranium, indicating the possibility of a group of elements analogous to the rare-earths.* In spite of the chemical similarity of uranium and plutonium, a satisfactory method for the separation of these elements, based on the different properties of the latter in the Pu^{IV} and Pu^{VI} oxidation states, has been developed.

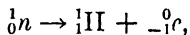
The long life of the element plutonium and the probability that upon neutron bombardment it would undergo fission, with the formation of large amounts of energy and the release of several additional neutrons, aroused interest in the possibility of its use as an exceptionally powerful explosive.

* Confirmed by the properties of elements 95 and 96 obtained by bombardment of ^{238}U and ^{239}Pu , respectively, by 40×10^6 e.v. helium nuclei.

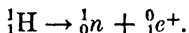
The preparation of plutonium has, therefore, been carried out in the following manner. A "pile" is erected, consisting of cylindrical slugs, or lumps, of pure uranium or uranium oxide, containing both the 235 and 238 isotopes, arranged in the form of a lattice in a structure of graphite. Upon exposure to a few neutrons, fission of some of the atoms occurs, with the resulting liberation of three or four fast neutrons for each atom that is split. These neutrons are partially slowed down by the graphite, so that some, with the "resonance" energies, are taken up by the ^{238}U atoms to form ^{239}U , which subsequently disintegrates to yield the fairly stable plutonium, while one of each group causes fission of a ^{235}U nucleus, thus yielding further neutrons, and so on. In place of graphite, heavy water can be used as a "moderator" for slowing down the fast neutrons produced in the fission process. By controlling the dimensions and spacing of the pile, and adjusting other factors, the formation of plutonium, once it has started, will continue steadily until either the ^{235}U is completely disintegrated or the equilibrium amount (cf. p. 131) of plutonium has been formed. Because of decreasing efficiency, the process is stopped before this stage is reached; the slugs are removed from the "pile," and the plutonium is separated from the uranium, and from the fission products, by chemical methods. The production of relatively large amounts in this manner of an entirely new element, which was not only unknown but which did not even exist a few years ago, represents one of the most spectacular achievements of modern science.*

In the so-called "atomic bomb," the essential requirement was the rapid liberation of large amounts of energy as the result of a self-propagating chain of fissions, utilizing fast neutrons, with the non-fission absorption being reduced to a minimum. It was desirable, therefore, to eliminate the ^{238}U isotope as far as possible; hence, ^{235}U , partially separated from the heavier isotope by one of the methods described earlier, and plutonium were employed as the fissionable materials. Calculations showed that a mass exceeding a certain critical size could be exploded by the presence of fast neutrons, which would initiate the rapidly branching chains. Subsequent events, which are now part of history, have served to confirm the theoretical predictions.³⁰

Nuclear Stability.—A detailed examination of the mass numbers and atomic numbers of stable and unstable isotopes may eventually throw light on the conditions determining nuclear stability. It has been seen already that there is a preference for an even number of neutrons, but in addition there appear to be limits for the ratio of protons to neutrons in stable nuclei. If as a result of an induced disintegration there is formed an unstable species containing too many neutrons, then electron emission will tend to occur; thus,



so that the neutron is replaced by a proton. On the other hand, if the proportion of protons is greater than required for stability a positron may be evolved; thus,



It is of interest to see how far the actual observations can be interpreted in the light of these considerations. Remembering that the number of

* Traces of ^{237}Np and ^{239}Pu may be present in the mineral pitchblende.

neutrons in a nucleus is equal to the difference in the mass number and atomic number, it can be readily deduced that the disintegrations (α ; p), (d ; p), (n ; α), (n ; p) and (n ; γ) lead to a relative increase in the number of neutrons and so the products should emit negatrons, whereas the interactions (α ; n), (d ; n), (n ; $2n$) and (γ ; n) result in a decrease in neutrons or gain of protons and hence the unstable nuclei may be expected to have positron activity. These anticipations are in agreement with the results already recorded. The (d ; α) change is accompanied by a loss of a neutron and a proton, and so it is not surprising to find that some products emit positrons whereas others send out negative electrons.

ISOTOPES AS INDICATORS

Isotopes as Tracers.—The chemical and physical properties of the isotopes of a given element may be regarded as identical, especially if the atomic weights are not too low. Since a radioactive isotope can be detected by its electron, positron or other activity, while one with an unusual mass can be identified in the mass spectrograph or in other ways, e.g., by density or heat capacity differences, it has been possible to use isotopes as "indicators" or "tracers" in various processes. The characteristic property of the isotope, viz., its radioactivity or its mass, acts as a "label" or "tag" which permits the fate of the element to be traced in many physical and chemical changes. Radioactive elements have the great advantage of easy detection, by means of a Geiger-Müller counter or by some form of electroscope; most artificial radioelements have relatively short lives, however, and so they cannot be used for experiments requiring an appreciable time. In such cases an isotope with an atomic weight which differs from that of the element in its common form has proved particularly valuable. Radioactive elements have been used to study the self-diffusion of metals, e.g., copper and gold, and of some salts; radio-carbon (^{14}C) has been employed to trace the fate of carbon in photosynthesis; radioactive isotopes of nitrogen (^{15}N) and of phosphorus (^{32}P) have been applied to follow the behavior of these elements in animal and plant metabolism. Other radioelements, such as sodium, potassium, iron, zinc and iodine, have also been used in biological absorption studies. A number of analytical applications of radioactive indicators have been described, particularly for the examination of sparingly soluble substances. Use has been made of both radioactive and nonradioactive isotopes for the investigation of reaction mechanisms and exchange processes. The nonradioactive deuterium and the ^{15}N isotope of nitrogen have been particularly important in connection with fatty acid and amino-acid metabolism studies. The availability of the relatively long lived hydrogen isotope, tritium, will make possible many interesting chemical and biological investigations.³¹

Some typical applications of isotopes as indicators are given below. In addition the use of indicators in adsorption and the determination of the surface areas of powders are described on page 1218.

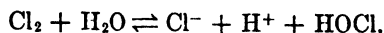
Analytical Applications.—The possibility of utilizing natural radioactive elements as indicators was realized by G. Hevesy and F. Paneth (1913), who showed that the solubilities of sparingly soluble salts could be determined by their aid. If a radioactive isotope is mixed with ordinary lead, then as a consequence of the virtual identity in physical and chemical properties of their compounds, the two isotopes will always remain in the same proportion irrespective of the physical state or chemical combination of the element. The active isotope should be one that does not decay too rapidly, for its concentration is estimated from its activity, and so radium-*D* is generally chosen for experiments with lead compounds, although the shorter-lived, but more active, thorium-*B* is also employed. A very minute amount of a radium-*D* compound, whose α -ray activity has been measured by an electroscope or other device, is mixed with a weighed quantity of an inactive lead compound, so that there is a known relationship between the radioactivity and the lead content of the system. The mixture is converted into a sparingly soluble salt, e.g., sulfate or chromate, and a saturated solution in water is prepared. A known volume is then evaporated to dryness and the radioactivity measured. Since the relationship between activity and amount of lead is known, the concentration of the latter in solution can be calculated. The solubility of ammonium phosphomolybdate in nitric acid has been determined in an analogous manner, using radio-phosphorus.

A somewhat similar method has been employed to determine the amount of lead remaining in solution after precipitation of the bulk for analytical purposes. A quantity of radium-*D*, of known activity, was added to a solution containing an unknown amount of lead; the bulk of the lead was then deposited electrolytically as the dioxide and weighed. The radioactivity of the residual solution was determined, by evaporating part or all to dryness, and the ratio of this to the original value gave the ratio of the amount of lead still remaining in solution to that precipitated (G. Hevesy, 1932).

Proof of the existence of hydrides of bismuth and lead was obtained by F. Paneth (1918, 1920) by means of radioactive materials. Magnesium filings were coated with thorium-*C* or radium-*C*, which are isotopic with bismuth, and dissolved in dilute hydrochloric acid. The gas evolved was found to be radioactive; it could be condensed in liquid air and decomposed by heat, and so probably contained bismuth hydride. By making a number of experiments of this kind the best conditions were discovered for the preparation of the hydride by the action of hydrochloric acid on an alloy of magnesium and ordinary bismuth. Evidence was obtained in a similar manner for the existence of the much less stable hydride of lead.³²

Exchange Reactions.—Information concerning the nature of chemical linkages and the mechanism of reactions has been obtained from a study of the exchange of isotopic atoms. If a solution of inactive lead chloride is mixed with one of lead nitrate containing a radioactive isotope of the metal, then on crystallizing out the lead chloride from the mixed solution it is found to possess activity; this is to be expected since both salts are ionized in solution and the lead ions do not belong to any particular anion. If a nonionizable compound, such as lead tetraphenyl, is used instead of the chloride, however, no exchange is observed and the organic compound remains inactive. Many experiments of this kind have been made with artificial radioelements; for example, exposure of sodium chloride to neutrons or deuterons yields some radio-chlorine,

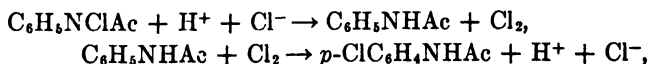
and if the salt is now dissolved in water and chlorine gas passed in, the resulting sodium chloride obtained on evaporation is less active than can be accounted for by the natural rate of decay of the radio-chlorine. There has evidently been an exchange between radio-chlorine ions and inactive chlorine molecules; some of the former have passed into the molecular form and been removed as gas, whereas some of the latter have become ions. An exchange of this kind can be readily understood because of the reaction



which occurs when chlorine is dissolved in water. Both radioactive and inactive chloride ions will take part in the equilibrium, and so some of the former will be converted into molecules. Similarly, free bromine and iodine have been shown to exchange with the corresponding ions in solution, presumably by an analogous mechanism. There is, however, as a general rule, no exchange between a halogen and an alkyl halide, e.g., iodine and ethyl or propyl iodide. By utilizing the radioactivity induced in sulfur, it has been possible to show that the sulfur atoms in sulfite and sulfate ions, and in sulfide and sulfate ions, do not exchange, but those in sulfide and thiosulfate can undergo interchange.

Reaction Mechanisms.—Several instances have been reported of information concerning reaction mechanisms obtained with the aid of artificial radioelements. When sodium iodide, previously activated by neutrons, is added to *sec*-octyl iodide in acetone solution exchange between the iodide ion and the organic iodine occurs; the rate can be determined by extracting the organic compound from time to time and measuring the radioactivity of the iodine contained in it, as well as that remaining in the solution. The interchange was found to occur at almost exactly the same rate, when expressed in the correct units, as the racemization of *d-sec*-octyl iodide by sodium iodide, thus showing conclusively that the stereochemical inversion reaction involves an exchange between iodide ions in solution and the iodine atom of the octyl iodide. Results of a similar character have been obtained for the racemization of α -phenyl ethyl bromide by bromide ions.

The isomeric change of N-chloroacetanilide to *p*-chloroacetanilide in the presence of hydrochloric acid has attracted much attention, and the mechanism proposed (K. J. P. Orton, 1910), in which chlorine is supposed to be formed as an intermediate, i.e.,



has received general support. The possibility has been considered, however, that the isomeric change may be intramolecular, occurring by exchange of valence bonds within a single molecule. Experiments have now been made on the reaction catalyzed by hydrochloric acid containing radioactive chloride ions, and the extent of the exchange determined. If the reaction were intramolecular, there is likely to be only a small amount of exchange between the chloride ions and the chlorine atom attached to the nitrogen in the original N-chloroacetanilide; on the other hand, if molecular chlorine is formed as an intermediate, considerable exchange should be observed, as will be evident from the equations given above. The results showed that, provided allowance is made for side reactions, the facts are in harmony with the latter mechanism.³³

Other applications of radioelements to the study of reaction mechanisms are given on pp. 1074 and 1142.

Deuterium as Indicator.—Deuterium has found considerable use in many directions; the proportion present in any mixture, or part of a system, containing hydrogen can be determined either by the mass spectrograph, or by methods based on the measurement of the heat conductivity of the gas, or of the density of the water produced by burning the hydrogen or compounds containing hydrogen. Many exchange reactions of hydrogen and deuterium, both in the gas phase and in solution, have been studied. All substances possessing hydroxyl groups, and so having a tendency to split off hydrogen ions, exchange readily with the deuterium of heavy water, as is to be expected. Further, the hydrogen atoms in acetylene, acetone and chloroform exchange with deuterium in alkaline solutions; although these substances are not to be regarded as appreciably ionized, there is undoubtedly a tendency in each case for hydrogen ions to be split off.

Study of reactions involving deuterium has provided information concerning the importance of zero-point energy (p. 1104), and has thrown light on various reactions. For example, the fact that hydrogen and deuterium react at equal rates with nitrous oxide shows that the essential step determining the velocity of the reaction involves hydrogen atoms and not molecules. If the latter were the case, the difference in zero-point energy would make the rates different, but the zero-point energies of the atoms are very small and would not influence the reaction velocity. Another use of deuterium is to compare the strengths of the bonds formed by hydrogen with oxygen, nitrogen and carbon; information may be obtained by a study of the rate of reaction between deuterium atoms and water vapor, ammonia and methane, respectively. Further aspects of the importance of deuterium in the study of reactions will be discussed in Chapter XIII.³⁴

REFERENCES

- Aston, F. W. "Mass Spectra and Isotopes."
 Feather, N. "Introduction to Nuclear Physics."
 Gamow, G. "Structure of Atomic Nuclei and Nuclear Transformations."
 Hahn, O. "Applied Radiochemistry."
 Hevesy, G. and Paneth, F. "A Manual of Radioactivity." (Translated by R. W. Lawson.)
 Paneth, F. "Radioelements as Indicators, etc."
 Pollard, E. C. and Davidson, W. L. "Applied Nuclear Physics."
 Rasetti, F. "Elements of Nuclear Physics."
 Rutherford (Lord), Chadwick, J. and Ellis, C. D. "Radiations from Radioactive Substances."
 Smyth, H. D. "Atomic Energy, etc."
 Soddy, F. "Interpretation of the atom."
 Stranathan, J. D. "The Particles of Modern Physics."
 Taylor, H. S. and Glasstone, S. "Treatise on Physical Chemistry," 3rd. ed., Vol. I, Chap. I (H. S. Taylor).
1. Becquerel, *Compt. rend.*, 122, 420, 501 (1896); P. and M. Curie, *ibid.*, 127, 175, 1215 (1898); Curie, *ibid.*, 130, 73, 77 (1900).
 2. Rutherford, *Phil. Mag.*, 12, 348, 13, 110 (1906); Rutherford and Geiger, *Proc. Roy. Soc.*, 81A, 141, 162 (1908); Crookes, *Chem. News*, 87, 241 (1903); Regener, *Verh. Deut. Phys. Ges.*, 10, 78 (1908).
 3. Ramsay and Soddy, *Proc. Roy. Soc.*, 72A, 204 (1903), 73A, 346; Rutherford and Royds, *Phil. Mag.*, 17, 281 (1909); Allen, *Phys. Rev.*, 61, 69 (1942).

4. Becquerel, *Compt. rend.*, **133**, 977 (1901); Rutherford and Soddy, *J. Chem. Soc.*, **81**, 837 (1902).
 5. Rutherford and Soddy, *Phil. Mag.*, **5**, 576 (1903); Ramsay and Whytlaw-Gray, *Proc. Roy. Soc.*, **84A**, 536 (1910).
 6. Soddy, "The Chemistry of the Radio-elements."
 7. Soddy, *Ann. Rep. Chem. Soc.*, **7**, 285 (1910); **10**, 262 (1913); Fleck, *J. Chem. Soc.*, **103**, 381, 1052 (1913); Fajans, *Phys. Z.*, **14**, 131, 136 (1913); Russell, *Chem. News*, **107**, 49 (1913); Soddy, *ibid.*, p. 97.
 8. Soddy and Hyman, *J. Chem. Soc.*, **105**, 1402 (1914); Richards and Lambert, *J. Am. Chem. Soc.*, **36**, 1329 (1914); Richards and Wadsworth, *ibid.*, **38**, 2613 (1916); Honigschmid, *Z. Elek.*, **23**, 161 (1917); **25**, 91 (1919); *Ber.*, **56**, 1837 (1923); *Z. anorg. Chem.*, **214**, 104 (1933); Baxter, *et al.*, *J. Am. Chem. Soc.*, **59**, 702 (1937); **60**, 62 (1938).
 9. Thomson, *Phil. Mag.*, **21**, 225 (1911); **24**, 209 (1912).
 10. Aston, *Proc. Roy. Soc.*, **89A**, 440 (1914); *Phil. Mag.*, **38**, 707 (1919); **39**, 449 (1920); **42**, 436 (1921); *Proc. Roy. Soc.*, **115A**, 487 (1927).
 11. Dempster, *Phys. Rev.*, **11**, 316 (1918); **18**, 415 (1921); **20**, 631 (1922).
 12. Bainbridge, *ibid.*, **39**, 1021; **42**, 1 (1932); *J. Frank. Inst.*, **215**, 509 (1933).
 13. Mattauch, *Phys. Z.*, **35**, 567 (1934); *Phys. Rev.*, **50**, 617, 1089 (1936); Dempster, *ibid.*, **50**, 98 (1936); Bainbridge and Jordan, *ibid.*, p. 282; **51**, 384 (1937); Bleakney and Hipple, *ibid.*, **53**, 521 (1938); Nier, *Rev. Sci. Instr.*, **11**, 212 (1940); Jordan, *et al.*, *J. App. Phys.*, **13**, 526, 539 (1942).
 14. Urey, Brickwedde and Murphy, *Phys. Rev.*, **39**, 164 (1932); **40**, 1 (1932); Giauque and Johnston, *J. Am. Chem. Soc.*, **51**, 1436 (1929); Birge and King, *Nature*, **124**, 127, 182 (1929).
 15. Aston, *Proc. Roy. Soc.*, **115A**, 487 (1927); "Mass Spectra and Isotopes"; Duckworth, *Phys. Rev.*, **62**, 19 (1942).
 16. Hertz, *Z. Phys.*, **79**, 108, 700 (1932), **91**, 810 (1934); Capron and de Hemptinne, *J. phys. Radium* [VII], **10**, 171 (1939); Clusius and Dickel, *Naturwiss.*, **26**, 546 (1938); *Z. phys. Chem.*, **B**, **44**, 397, 451 (1939); Taylor, *Nature*, **144**, 8 (1939); Nier and Bardeen, *J. Chem. Phys.*, **9**, 690 (1941).
 17. Brønsted and Hevesy, *Nature*, **106**, 144 (1920); *Phil. Mag.*, **43**, 31 (1922); Harkins, *J. Am. Chem. Soc.*, **43**, 1083 (1921); **45**, 591 (1923), **48**, 58 (1926); *Phil. Mag.*, **6**, 601 (1928); Mulliken, *J. Am. Chem. Soc.*, **44**, 2387 (1922); **45**, 1592 (1923).
 18. Oliphant, Shire and Crowther, *Proc. Roy. Soc.*, **146A**, 922 (1934); Smythe and Hemmendinger, *Phys. Rev.*, **51**, 178, 1052 (1937).
 19. Washburn and Urey, *Proc. Nat. Acad. Sci.*, **18**, 496 (1932); Lewis, *J. Am. Chem. Soc.*, **55**, 1297 (1933); Lewis and Macdonald, *J. Chem. Phys.*, **1**, 341 (1933); Taylor, Eyring and Frost, *ibid.*, p. 823; Farkas, "Orthohydrogen, Parahydrogen and Heavy Hydrogen"; Walton and Wolfenden, *Trans. Farad. Soc.*, **34**, 436 (1938); Tronstad and Brun, *ibid.*, p. 766; Johnston and Hutchison, *J. Chem. Phys.*, **8**, 869 (1940); **9**, 469 (1941).
 20. Urey, *et al.*, *J. Am. Chem. Soc.*, **59**, 1407 (1937); *Ind. Eng. Chem.*, **29**, 531 (1937); *J. Chem. Phys.*, **4**, 622 (1936); **5**, 856 (1937); **6**, 296 (1938); **7**, 34, 137 (1939); **8**, 532 (1940); Stewart and Cohen, *ibid.*, p. 904.
- For reviews on the separation of isotopes, see Walker, *Ann. Rep. Chem. Soc.*, **35**, 134 (1938); Urey, *Phys. Soc. Rep. Prog. Phys.*, **6**, 48 (1939); *J. App. Phys.*, **12**, 270 (1941); Taylor, *op. cit.*; Smyth, *op. cit.*
21. Farkas, *ref.* (19); Urey and Teal, *Rev. Mod. Phys.*, **7**, 34 (1935); see also, *Ann. Rep. Chem. Soc.*, 1933, 1934, 1935 and 1937; *Z. Elek.*, **44**, 1 *et seq.* (1938); Selwood, *J. Chem. Ed.*, **18**, 515 (1941).
 22. Rutherford and Chadwick, *Phil. Mag.*, **42**, 809 (1921); **44**, 417 (1922); for complete bibliography, see Livingston and Bethe, *Rev. Mod. Phys.*, **9**, 245 (1937).
 23. Feather, *Proc. Roy. Soc.*, **136A**, 709 (1932); Fermi, *et al.*, *ibid.*, **146A**, 483 (1934); **149A**, 522 (1935).

24. Cockroft and Walton, *Proc. Roy. Soc.*, **136A**, 619 (1932); **137A**, 229; Kirchner, *Phys. Z.*, **33**, 777 (1932); Lawrence and Livingston, *Phys. Rev.*, **40**, 19 (1932); **45**, 608 (1934); Lawrence and Cooksey, *ibid.*, **50**, 1131 (1936); Van de Graaff, *et al.*, *ibid.*, **43**, 158 (1933); **49**, 761 (1936); Lewis, Lawrence and Livingston, *ibid.*, **44**, 55, 317 (1933); Livingston and Bethe, ref. (22).
25. Bethe, *Phys. Rev.*, **47**, 633 (1935); **53**, 313 (1938); Oliphant, Kempton and Rutherford, *Proc. Roy. Soc.*, **150A**, 253 (1935).
26. Curie and Joliot, *Compt. rend.*, **198**, 254, 559 (1934); *Nature*, **133**, 201 (1934).
27. Fermi, *et al.*, ref. (23).
- For reviews of artificial radioactivity, see Livingston and Bethe, ref. (22); Ridenour and Yost, *Chem. Rev.*, **18**, 457 (1936); Seaborg, *ibid.*, **27**, 199 (1940); Livingood and Seaborg, *Rev. Mod. Phys.*, **12**, 230 (1940); Seaborg, *ibid.*, **16**, 1 (1944); Fleischmann, *Z. Elek.*, **47**, 8 (1941).
28. Weiszäcker, *Naturwiss.*, **24**, 813 (1936); Snell, *Phys. Rev.*, **52**, 1007 (1937); Mitchell and Langer, *ibid.*, **53**, 505 (1938); Segrè, Halford and Seaborg, *ibid.*, **55**, 321 (1939); Seaborg and Segrè, *ibid.*, p. 808; DeVault and Libby, *ibid.*, **58**, 688 (1940); *J. Am. Chem. Soc.*, **63**, 3216 (1941); Sugden, *et al.*, *Nature*, **143**, 517 (1939); Pontecorvo, *ibid.*, **144**, 212 (1939); Willard, *J. Am. Chem. Soc.*, **62**, 3161 (1940); **64**, 1342 (1942); Cooper, *Phys. Rev.*, **61**, 1 (1942).
29. Perrier and Segrè, *J. Chem. Phys.*, **5**, 715 (1937), **7**, 155 (1939); Alvarez and Cornog, *Phys. Rev.*, **56**, 613 (1939); Segrè, *et al.*, *ibid.*, **57**, 1087 (1940); **58**, 672 (1940); *Sci. Monthly*, **57**, 12 (1943); Irvine, *Chem. Rev.*, **23**, 77 (1938); Paneth, *Nature*, **149**, 565 (1942).
30. Hahn and Strassmann, *Naturwiss.*, **27**, 11, 89 (1939); **30**, 256 (1942); Meitner and Frisch, *Nature*, **143**, 240 (1939); Feather, *ibid.*, p. 877; Turner, *Rev. Mod. Phys.*, **12**, 1 (1940); Nishina, *et al.*, *Phys. Rev.*, **58**, 660 (1940), **59**, 323, 677 (1941); Nier, *et al.*, *ibid.*, **57**, 546, 748 (1940); Kingdon, *et al.*, p. 749; McMillan and Abelson, *ibid.*, p. 1185; Smyth, *op. cit.*; *Rev. Mod. Phys.*, **17**, 351 (1945).
31. For reviews, see Geib, *Angew. Chem.*, **51**, 622 (1938); Enders, *ibid.*, **53**, 28 (1940); Reitz, *Z. Elek.*, **45**, 100 (1939); Rosenblum and Flagg, *J. Frank. Inst.*, **228**, 471, 623 (1939); Rosenblum, *J. Chem. Ed.*, **17**, 567 (1940); Brewer, *ibid.*, **18**, 217 (1941); Timm, *ibid.*, **20**, 54 (1943); Loofburow, *Rev. Mod. Phys.*, **12**, 217 (1940); Urey, *J. App. Phys.*, **12**, 270 (1941); Rittenberg, *ibid.*, **13**, 561 (1942); Schönheimer and Rittenberg, *Physiol. Rev.*, **20**, 218 (1940); Hevesy, *Ann. Rev. Biochem.*, **9**, 641 (1940); Schönheimer and Ratner, *ibid.*, **10**, 197 (1941); Kurbatov and Pool, *Chem. Rev.*, **32**, 231 (1943).
32. Hevesy and Paneth, *Z. anorg. Chem.*, **82**, 323 (1913); Paneth, *ibid.*, **35**, 549 (1922); *Z. Elek.*, **24**, 298 (1918); *Ber.*, **51**, 1704 (1918); *Angew. Chem.*, **35**, 549 (1922); "Radioelements as Indicators."
33. Topley, Weiss, *et al.*, *J. Chem. Soc.*, 1295, 1525 (1935); 912, 1173 (1936); Olson, *J. Am. Chem. Soc.*, **58**, 2467 (1936); **59**, 1613 (1937).
34. For reviews, see references under (21).

CHAPTER III

THE FIRST AND SECOND LAWS OF THERMODYNAMICS

THE FIRST LAW OF THERMODYNAMICS

Energy.—The energy of a body may be defined broadly as its capacity for doing work; this energy may take various forms, e.g., kinetic energy of a body in motion, potential energy due to position, heat energy, chemical energy, electrical energy, surface energy, etc. Chemical and physical changes are almost invariably accompanied by energy changes, and results of great importance have been reached by studying the laws underlying these changes; this study constitutes the subject matter of thermodynamics. For physical chemistry, the energy possessed by a body by virtue of its position in space is not so important as the energy of the atoms and molecules of which the body is composed. This is made up of the translational kinetic energy of the moving molecule, of the energies of vibration and rotation within the molecule, of the internal potential energy determined by the arrangement of the nuclei and electrons, and any other forms of energy involved in the structure of matter. From the standpoint of thermodynamics, however, it is not necessary to know the actual value of the energy, but rather the change accompanying a particular process.

↳ **Energy Units.**—All forms of energy have the dimensions m^2t^{-2} , that is, mass \times (length)²/(time)²; and the units used in scientific work are almost invariably the gram (g.), centimeter (cm.) and second(sec.); these form the basis of the c.g.s. system. The meter, or 100 cm., was originally defined as 10^{-7} times the length of the earth's quadrant, from north pole to equator, passing through Paris. The standard meter is, however, the distance between two marks on a bar at 0° c., kept at Sèvres, near Paris; the earth's quadrant is actually 1.00021×10^7 meters. The standard kilogram (1000 g.) is taken as the mass of a lump of platinum, also at Sèvres; it was intended to be equal to the weight of 1000 cc. of "pure" water at its temperature of maximum density, 4° c., but actually there is a small difference. The liter is the volume occupied by 1 kilogram of water at 4° c. and 1 atm. pressure, and was meant to be 1000 cc., but it is actually 1000.027 cc.; because of this discrepancy it is becoming the practice to express volumes of liquids and gases in terms of the milliliter (ml.), i.e., a thousandth part of a liter, instead of in cc. The second is defined as 1/86,400 part of a mean solar day, i.e., the average interval between successive transits of the sun across the meridian at any given place. The unit of energy on the c.g.s. system is the erg; it is the work done when a force of 1 dyne acts through a distance of 1 cm., the dyne being the force which acting for 1 sec. on a mass of 1 g. produces in it a velocity of motion of 1 cm. per sec. Since the erg is so small, a subsidiary unit, the joule, is defined as 10^7 ergs. Unfortunately

confusion has arisen because of the necessity for distinguishing between the absolute joule, as thus defined, and the international joule, based on electrical measurements. Theoretically, the absolute volt-coulomb should be equal to 10^7 ergs, i.e., 1 joule (see p. 6), but the actual definition of the volt, in terms of the Weston cell, and of the coulomb, in terms of the quantity of silver deposited (p. 885), makes the international volt-coulomb, or int. joule, equal to 1.0002 abs. joules. It is necessary, therefore, to distinguish between int. joules and abs. joules; there is no doubt, of course, concerning the erg, and this should be used if possible to avoid confusion.

Equivalence of Work and Heat.—The relationship between mechanical work and heat was first clearly seen by Count B. Rumford (1798) who studied the rise of temperature which accompanied the boring of a cannon; he concluded that the mechanical work involved in the boring was responsible for the heat produced. Shortly afterward, H. Davy (1799) showed that by rubbing two pieces of ice together melting occurred, the latent heat of fusion being obtained from the work done against friction. The subject did not attract any great interest until 1840 when J. P. Joule commenced his classical experiments, lasting nearly forty years, on the relationship between work expended and the heat produced. Before any of Joule's results were published, however, J. R. Mayer (1842) enunciated a general relationship concerning the equivalence of heat and mechanical work, and made the first approximate evaluation of what is now known as the **mechanical equivalent of heat**, that is, the ratio of the work done to the corresponding heat change. Mayer's calculation was based on a knowledge of specific heats and of the work done by a gas in expansion; these topics will, therefore, be considered next.*

Specific Heats: Constant Pressure and Constant Volume.—The **specific heat** is defined as the quantity of heat required to raise the temperature of 1 gram of a substance by 1° , the value being expressed in **calories**, where the "15° calorie" is the amount necessary to heat 1 gram of water from 14.5° to 15.5° C.† Although specific heats generally vary with temperature, this factor will be disregarded for the present because of others of greater importance. For all substances, and particularly for gases, the quantity of heat required to raise the temperature by a given amount depends on the possibility of variation, or otherwise, of pressure and volume; in order to obtain definite results it is necessary to specify the conditions. Two forms of the specific heat are, therefore, in common use; they are obtained either by keeping the volume constant and allowing the pressure to increase, or else by maintaining the pressure constant and

* Danish writers claim that L. A. Colding (1842) should have priority over Mayer in connection with the relationship of work and heat.

† Because heat measurements are usually standardized by reference to an amount, expressed in int. joules, obtained by the passage of a known quantity of electricity through a resistance, the U. S. Bureau of Standards recommends that a calorie be defined as the heat equivalent of 4.1833 int. joules, i.e., 4.1841 abs. joules. This defined calorie differs slightly from the standard 15° calorie; it corresponds more closely to the specific heat of water at 17° C.

permitting the volume to increase when the temperature is raised. The values so obtained are the specific heats at constant volume (c_v) or at constant pressure (c_p), respectively. For gases, the difference between the two specific heats is quite appreciable, the latter being the larger; this is due to the fact that when the gas is heated at constant pressure, expansion occurs and consequently work has to be done against the pressure of the atmosphere. The energy necessary to perform this work must evidently be supplied in the form of heat, so that the difference in the two specific heats is equal to the work done in expansion when the temperature of 1 g. of substance is raised by 1° at constant (atmospheric) pressure, provided there is no accompanying change in the internal energy.

Work Done in Expansion.—The work done by a system when it expands against a constant external pressure may be evaluated in the following manner. Consider any substance contained in a cylinder, of cross section a sq. cm., fitted with a piston * upon which is exerted a constant pressure of P dynes per sq. cm. The total force acting on the piston is thus Pa dynes. Suppose that when the substance in the cylinder expands, the external pressure remaining constant, the piston is raised through a height h cm. The work done against the external pressure is equal to the product of the force acting and the displacement of the point of application, and hence is $Pa \times h$ ergs. The quantity ah is equal to ΔV , the increase in volume during the expansion, and consequently

$$\text{Work done in expansion} = P\Delta V. \quad (1)$$

The work done in an expansion against a constant external pressure is thus equal to the product of the pressure and the increase of volume.

If the external pressure does not remain constant during the whole expansion process, it is necessary to consider a series of very small volume changes dV , in each of which the work done is PdV , where P is the pressure at that instant. The total work done is then the integral of PdV over the path followed in the expansion process.

The Mechanical Equivalent of Heat.—If the internal energy of a gas is independent of its volume (see p. 191), as was tacitly assumed by J. R. Mayer (1842), the difference between the specific heats at constant pressure and constant volume may be regarded as exactly equivalent to the work done when the gas expands at constant pressure. At the time Mayer made this calculation the accepted values for the specific heats of air were $c_p = 0.267$ cal., and $c_v = 0.1875$ cal., so that the difference is 0.0795 cal., and this should be equivalent to the mechanical work done in expansion when the temperature of 1 g. of air is raised by 1° . The volume of 1 g. of air at 0° c. and 1 atm. pressure was taken as 770 cc., and according to Gay-Lussac's law (p. 192) the increase for 1° rise of temperature should be $770/273$ cc.; this is, consequently, the value of ΔV in cc. The pressures of the atmosphere, 76 cm. of mercury, expressed in

* In all thermodynamic calculations it is assumed that the piston is weightless and moves without friction, so that no work is lost in the actual movement of the piston itself.

c.g.s. units is $76 \times 981 \times 13.6$ dynes per sq. cm.,* and so the work done in expansion ($P\Delta V$) is

$$\frac{76 \times 981 \times 13.6 \times 770}{273} = 2.86 \times 10^6 \text{ ergs.}$$

Equating this mechanical work to the heat supplied, 0.0795 cal., it follows that

$$1 \text{ calorie} \equiv \frac{2.86 \times 10^6}{0.0795} = 3.6 \times 10^7 \text{ ergs.}$$

This is the result, expressed in modern units, as obtained by Mayer, but if more recent data are used in the calculation, thus $c_p - c_v = 0.0687$ cal., and the volume of 1 g. of air at 0°C. and 1 atm. pressure = 773.4 cc., it is found that

$$1 \text{ calorie} \equiv 4.18 \times 10^7 \text{ ergs.}$$

Although Mayer's calculation represented the first estimate of the mechanical equivalent of heat, it is to J. P. Joule that credit must be given for a complete proof, published between 1843 and 1878, of the equivalence of heat energy and mechanical work. He used a variety of ways of converting work into heat, and found that in every case the production of 1 calorie required the expenditure of a constant amount of energy, within the limits of experimental error. Among the methods adopted by Joule may be mentioned the following: heat produced by stirring water or mercury by paddle wheels; heat evolved or absorbed when air was compressed or expanded; heat generated when water was forced through capillaries; heat produced by electric currents induced in a coil of wire when rotated between the poles of a magnet; and heat generated by the passage of a definite quantity of electricity through a coil of known resistance. From Joule's most accurate experiments the mechanical equivalent of heat was found to be 4.17×10^7 ergs per calorie, independent of the method of conversion; the constancy of this result went far toward establishing the generalization that the expenditure of a certain amount of work was accompanied by a definite amount of heat. Later workers have generally found the method involving electrical heating to be the most accurate, and the accepted value of the mechanical equivalent (see p. 182, footnote) is given by

$$\begin{aligned} 1 \text{ (15}^\circ\text{) calorie} &\equiv 4.1858 \times 10^7 \text{ ergs} \equiv 4.1858 \text{ abs. joules} \\ 1 \text{ (defined) calorie} &\equiv 4.1833 \text{ int. joules} \equiv 4.1841 \text{ abs. joules} \\ 1 \text{ joule} &\equiv 0.2390 \text{ (defined) cal.} \equiv 0.2389 \text{ (15}^\circ\text{) cal.} \end{aligned}$$

Law of Conservation of Energy: First Law of Thermodynamics.—The impossibility of perpetual motion, that is, the continuous production of mechanical work without supplying an equivalent amount of energy from another source, was widely accepted even at the beginning of the 18th century, and in 1847 H. von Helmholtz showed that this view and the equivalence of work and heat were aspects of a wider generalization which has become known as the **law of conservation of energy**. This law has been enunciated in various forms, but its fundamental implication

* The hydrostatic pressure of any column of fluid, liquid or gas, of height h cm. is equal to $h\rho g$ dynes per sq. cm., where g is the acceleration due to gravity, in cm. per sec. per sec., and ρ is the density of the fluid; for most purposes g may be taken as 981.

is that *although energy may be converted from one form to another it cannot be created or destroyed*: in other words, *whenever a quantity of one kind of energy disappears an exactly equivalent amount of other kinds must be produced*.^{*} It is evident that perpetual motion would be contrary to this principle, for it would involve the creation of energy. Further, the exact equivalence of mechanical work and heat, independent of the manner whereby one is converted into the other, as found by Joule and others, is a necessary consequence of the law of conservation of energy. If this were not so, it would be possible to convert work into heat by one method and then to reverse the process by a method giving a larger amount of work for the same quantity of heat. The net result would be the production of mechanical work although no other form of energy had been used up; this would, however, be contrary to the law. It must be emphasized that the law of conservation of energy is purely the result of experience, namely the impossibility of any form of perpetual motion, and it is on the assumption that this experience is universal that the **first law of thermodynamics** is based. The law can be stated in any of the ways given above for the principle of the conservation of energy, or else in the following form: *The total energy of a system † and its surroundings must remain constant, although it may be changed from one form to another.*

Applications of the First Law:

Energy of a System.—Consider any system, viz., gas, liquid or solid, represented by the state *A* in Fig. 34, where the coordinates are properties, e.g., pressure and volume, on which the energy of the system depends; suppose the conditions are now altered so that the system moves to *B* by the path I, and is then brought back to the state *A* by a different path II. It is a consequence of the first law of thermodynamics that the total energy change in path I must be identical with that in path II, for if it were not, perpetual motion would be possible. Imagine, for example, that the increase in energy involved in

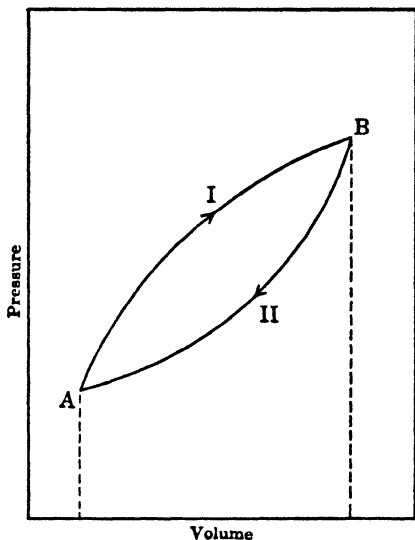


Fig. 34. Variation of pressure and volume

^{*} According to the concept of the equivalence of mass and energy, based on the theory of relativity (p. 7), mass is to be regarded as a form of energy. The laws of conservation of mass and energy are thus complementary.

† The term "system" is used in thermodynamics to refer to a specified portion of matter containing a definite amount (or amounts) of a given substance (or substances). The combination of the system and its surroundings cannot exchange energy with any other system and its surroundings, and so is sometimes called an "isolated system."

path I were greater than the decrease in returning by II; then by carrying out the processes $A \rightarrow B$ by path I, and $B \rightarrow A$ by path II, the system would be brought back to its initial state but there would be a residuum of energy; in other words, energy would have been created. Since this is contrary to the first law, it must be concluded that the net energy change of a system when passing from a state A to a state B must depend only on the initial and final states and must be independent of the path followed. It is, therefore, possible to ascribe a definite amount of energy to any given state, which is determined only by the state itself and not by its previous history. If E_A represents the energy in the state A and E_B that in the state B , then the increase in energy ΔE in passing from A to B is given by

$$\{\Delta E = E_B - E_A \quad (2)$$

and is independent of the path taken. The quantity E is frequently called the **internal** or **intrinsic energy**, but it will be sufficient to refer to it merely as the energy or energy content of the system; it includes all forms of energy other than those resulting from the position of the system in space, for this is assumed to remain constant. Although the actual value of the energy is not known, it does not matter, for, as already indicated, thermodynamics is concerned with *changes* in the energy, and these are measurable quantities.

The energy of a system is determined not only by the state of the system, but also by the amount (or amounts) of the substance (or substances) contained within it. If the quantity of material in the system is changed the energy will be changed proportionately. This is an example of an **extensive property**. Another familiar illustration of such a property is volume. Extensive properties may be distinguished from **intensive properties**, such as density, refractive index, vapor pressure, etc., which are independent of the quantity of material concerned.

Complete (or Exact) Differentials.—Consider an extensive property X of a system which is a single-valued function of the variables, such as temperature, pressure, volume, and masses of the constituents, that determine the state of the system; it is then possible to write

$$X = f(x, y, z, \dots),$$

where x, y, z, \dots , represent the variables which completely determine the value of X for the system. If the system is changed from the state A to the state B , the increase ΔX in the property X is defined exactly by

$$X_B - X_A = \Delta X = f(x_B, y_B, z_B, \dots) - f(x_A, y_A, z_A, \dots),$$

so that it is determined solely by the initial and final states of the system. Under these conditions a small change dX can be expressed in the form

$$dX = \frac{\partial X}{\partial x} dx + \frac{\partial X}{\partial y} dy + \frac{\partial X}{\partial z} dz + \dots, \quad (3)$$

the use of the partial differential notation, e.g., $\partial X/\partial x$, implying that all variables other than x are assumed constant for that particular term. The foregoing properties of X are often summarized in the statement that dX is a **complete** or **exact differential**; the change in X accompanying a change from a state A to a state B is then dependent only upon these states, and is independent of the means whereby the change has been brought about. As seen above, the energy E of a system is just such a property, and hence dE is a complete differential. The energy of a given mass is dependent on the pressure, volume and temperature, but since these three quantities are related by some form of equation of state (see Chapter IV), only two of them are independent, e.g., volume and temperature or pressure and temperature. It is then possible to write (3) in the alternative forms

$$dE = \left(\frac{\partial E}{\partial V} \right)_T dV + \left(\frac{\partial E}{\partial T} \right)_V dT, \quad (4a)$$

$$dE = \left(\frac{\partial E}{\partial P} \right)_T dP + \left(\frac{\partial E}{\partial T} \right)_P dT. \quad (4b)$$

Applications of these and analogous equations will be given later.

Energy, Work and Heat Changes.—When a system changes from one state to another, it may lose or gain energy as heat and work. The work, represented by the symbol W , where W is positive, by convention, when work is done *by* the system and negative when work is done *on* the system, may be mechanical work of expansion, electrical work, etc. The heat absorbed by the system is represented by Q , so that $-Q$ would be the heat given up by it. If, in a change from the state A to the state B , the energy content of a system is increased by ΔE , the work done being W and the heat absorbed Q , then by the first law of thermodynamics

$$\Delta E = Q - W, \quad (5)$$

in order that the total energy of the system and its surroundings shall remain unchanged in passing from A to B .

For an infinitesimal stage of an appreciable change, (5) is sometimes written in the form

$$dE = dQ - dW, \quad (5a)$$

but this expression is liable to be misleading. Although E is a definite property of the system, and dE is a complete differential, Q and W are not definite properties, and dQ and dW are not complete differentials. The values of Q and W associated with a particular change of state depend on the particular path followed in the change. It was indicated earlier that the work done against the external pressure in any process, e.g., from A to B in Fig. 34, is equal to the integral of the PdV terms over the path followed from A to B ; this is equivalent to the area under

the curve representing the change on a pressure-volume diagram.* It is at once evident from an examination of Fig. 34, that the work done in passing from A to B depends on the nature of the path followed. Since Q , W and ΔE are related by (5), and ΔE is independent of the path, it is obvious that Q , like W , must vary with the path taken in passing from state A to state B . It follows, therefore, that small changes in the heat and work terms cannot be regarded as complete differentials.

In order to avoid the misunderstanding which may arise from the use of the symbols dQ and dW , it is preferable to write (5a) as

$$dE = q - w, \quad (6)$$

where dE is the increase in the energy content of the system accompanying an infinitesimal stage of an appreciable change; q is the small quantity of heat absorbed and w is the work done by the system at the same time. The algebraic sums of the q and of the w terms for the whole process give Q and W , respectively.

There are certain conditions under which the heat and work terms become definite; two of these are constant pressure and constant volume, respectively, provided the work is only mechanical work done in expansion (or compression) against an external pressure. At constant pressure the work done is equal to $P\Delta V$, and since the volume change is a definite quantity, independent of the path, it is evident that W will also be definite. It is then possible to write (5) as

$$\Delta E = Q_P - P\Delta V \quad \text{or} \quad Q_P = \Delta E + P\Delta V, \quad (7)$$

the subscript P being used to indicate heat absorbed at constant pressure. Since ΔE and $P\Delta V$ are determined only by the initial and final states of the system, it follows that the heat absorbed at constant pressure is also independent of the path.

At constant volume ($\Delta V = 0$), on the other hand, no external work is done, so that W is zero; (5) then becomes

$$Q_V = \Delta E = E_B - E_A, \quad (8)$$

so that the heat absorbed in a constant volume process is equal to the increase of the energy of the system. The latter depends only on the initial and final states, and hence the same must apply to the heat change at constant volume.

The Heat Content.—If ΔE is replaced by $E_B - E_A$, and the increase of volume ΔV is represented by $V_B - V_A$, then (7) becomes

$$\begin{aligned} Q_P &= (E_B - E_A) + P(V_B - V_A) \\ &= (E_B + PV_B) - (E_A + PV_A). \end{aligned} \quad (9)$$

Since P and V are properties of the state of the system, it is evident that the quantity $E + PV$, like the energy E , is dependent only on the state

* It is assumed that the external pressure P is always virtually identical with the pressure of the gas (see p. 196).

and not on the previous history of the system. The symbol H is used to represent $E + PV$, and it is generally called the **heat content** of the system, although the terms **heat function at constant pressure** and **enthalpy** have been employed. Rewriting (9) with this new symbol, it is seen that

$$Q_P = H_B - H_A = \Delta H, \quad (10)$$

so that the increase of heat content of a system is equal to the heat absorbed at constant pressure. Substituting the value ΔH for Q_P in (7) gives

$$\Delta H = \Delta E + P\Delta V, \quad (11)$$

where ΔE refers to constant volume and ΔH to constant pressure conditions.

Heat Capacity.—The **heat capacity** of a system is defined as the quantity of heat required to raise the temperature by 1° . If the system consists of a single substance, or a solution, and its weight is 1 g., then the heat capacity is called the **specific heat capacity**, or the **specific heat**, as already defined (p. 182). Similarly, for 1 mole of a pure substance the heat capacity of the system is the **molar heat capacity**, or the **molar heat**, of the substance concerned; it is equal to the product of the molecular weight and the specific heat. ✓

Since the heat capacity C may vary with the temperature, it is preferable to define it in the form

$$C = q/dT, \quad (12)$$

where q may be regarded as the infinitesimally small amount of heat absorbed by the system when the temperature is raised by dT degrees. Since q is, in general, an indefinite quantity, the heat capacity will be indefinite unless certain conditions, particularly constant volume or constant pressure, as stated above, are specified. The heat capacity at constant volume C_V is then given by (12) as

$$C_V = q_V/dT. \quad (13)$$

According to (8), the heat Q_V absorbed at constant volume is equal to the increase ΔE in the energy E under the same conditions; for a small change in the system the corresponding quantities are q_V and dE_V , and so the heat capacity at constant volume, as defined by (13), may be written as

$$C_V = \left(\frac{\partial E}{\partial T} \right)_V. \quad (14)$$

The heat capacity at constant volume is thus equal to the rate of change of the energy of the system with temperature, at constant volume.

At constant pressure, the heat capacity C_P is represented by

$$C_P = q_P/dT, \quad (15)$$

and since q_P , the heat absorbed, is equal to dH at constant pressure, by

(10), it follows that

$$C_P = \left(\frac{\partial H}{\partial T} \right)_P. \quad (16)$$

Ideal Gases.—As already seen, since dE is a complete differential, it is possible to write for a definite mass of any substance, e.g., a gas,

$$dE = \left(\frac{\partial E}{\partial V} \right)_T dV + \left(\frac{\partial E}{\partial T} \right)_V dT. \quad (17)$$

By (14), $(\partial E / \partial T)_V$ is equal to C_V ; hence (17) becomes

$$dE = \left(\frac{\partial E}{\partial V} \right)_T dV + C_V dT. \quad (18)$$

For an infinitesimal stage of a process, the fundamental equation (6) may be written in the form

$$q = dE + w,$$

and introduction of (18) gives

$$q = \left(\frac{\partial E}{\partial V} \right)_T dV + C_V dT + w, \quad (19)$$

where w may be taken as the work done against the external pressure.

According to experiments made by J. L. Gay-Lussac (1807) and J. P. Joule (1844), when a gas expands into a vacuum, thus doing no external work, there is no loss or gain of heat. Two similar copper vessels, one containing air under pressure and the other evacuated, were connected by a tube with a stopcock; when this was opened the temperature of the vessel containing the air fell, but that of the one originally evacuated rose by an equal amount. It appeared, therefore, that there was no net heat change when the volume of the gas was increased, provided no external work was done. Subsequent experiments, carried out along somewhat different lines, by J. P. Joule in conjunction with W. Thomson (Lord Kelvin), between 1852 and 1862, showed that there is actually a small temperature change when a gas expands freely without doing external work. This heat change is related to the Joule-Thomson effect, which is considered in Chapter IV where it will be shown that it is connected with the size of the molecules and the attractive forces existing between them. {An ideal gas, however, may be regarded as one for which the molecular attraction is negligibly small and in which the actual volume of the molecules is small in comparison with the space they inhabit, so that the Joule-Thomson effect should be zero.} For an ideal gas, therefore, there should be no absorption or evolution of heat upon expansion, provided no external work is done. It is seen, therefore, that if w is zero, q is also zero for an ideal gas, and hence at constant temperature, when

dT is zero, it follows from (19) that

$$\left(\frac{\partial E}{\partial V}\right)_T dV = 0. \quad (20)$$

In the case under consideration, the gas originally occupies one vessel and expands to fill two, and so dV is not zero; consequently for an ideal gas

$$\left(\frac{\partial E}{\partial V}\right)_T = 0. \quad (21)$$

It is to be concluded, therefore, that at constant temperature the energy of an ideal gas is independent of its volume; this condition, which should apply at all temperatures, is sometimes used to define an ideal gas.

THERMODYNAMICS OF AN IDEAL GAS

✓ **Boyle's Law and Gay-Lussac's Law.**—For ordinary purposes an alternative, more direct, method is used to describe an ideal gas; it may be defined as a gas to which the laws of Boyle, Gay-Lussac, and Avogadro are applicable at all temperatures. The conclusions reached by R. Boyle (1662), from his experiments on the compressibility of air, have been generalized in the statement known as **Boyle's law**,* as follows. *At constant temperature the volume of a definite mass of gas is inversely proportional to the pressure; that is,*

$$V \propto 1/P \quad \text{or} \quad PV = \text{constant}. \quad (21)$$

It will be seen in Chapter IV that the behavior of actual gases departs appreciably from the requirements of this equation, but it can be deduced from the kinetic theory (p. 251), provided the molecules can be regarded as point masses having no attraction for one another, that the product PV should be a constant at each temperature. The applicability of Boyle's law may thus be regarded as one of the criteria of an ideal gas.

The variation of the volume of a gas with temperature, at constant pressure, is expressed by the generalization often called **Gay-Lussac's law**; actually, analogous conclusions, which remained unpublished, had been reached by J. A. C. Charles (1787)—hence the alternative name, Charles's law—before J. L. Gay-Lussac reported his results in 1802. In its simplest form the law states that *at constant pressure, the volume of any gas expands by the same fraction of its volume at 0°C. for every 1° rise of temperature*; thus, if V_0 is the volume of a given mass of gas at 0°C. and V_t is that at $t^\circ \text{C.}$, then $V_t - V_0 = \alpha_v t V_0$, or

$$V_t = V_0(1 + \alpha_v t), \quad (22)$$

where α_v , the coefficient of cubical expansion, should be the same for all

* On the continent of Europe it is sometimes called Mariotte's law, after E. Mariotte who, in 1676, obtained results similar to those of Boyle.

gases. In actual practice, this equation is not obeyed exactly (p. 247), but at low pressures and at relatively high temperatures, that is, as the gas approaches ideal behavior, the coefficients of expansion (α_v) of all gases tend toward a constant figure. The most recent observations (W. H. Keesom, 1934) indicate that the limiting value is 0.0036609, and hence this may be taken as α_v for an ideal gas, applicable at all temperatures and pressures. For such a gas, therefore, (22) becomes

$$V_t = V_0(1 + 0.0036609t); \quad (23)$$

$$\therefore V_t = V_0 \left(1 + \frac{t}{273.16} \right) \quad \text{or} \quad V_t = \frac{V_0}{273.16} (273.16 + t). \quad (24)$$

If this relationship remains applicable at very low temperatures, it is evident that the volume of an ideal gas must become zero when the temperature is -273.16°C. ; this is known as the **absolute zero** of the ideal gas scale of temperature. The addition of 273.16° to the Centigrade temperature gives the so-called **absolute scale**, so that 0°C. is 273.16° Absolute, represented by T_0 , and $t^\circ \text{C.}$ is $273.16 + t^\circ$ Absolute, represented by T_t ; (24) may, therefore, be written

$$V_t = \frac{V_0}{T_0} T_t \quad \text{or} \quad \frac{V_t}{T_t} = \frac{V_0}{T_0}. \quad (25)$$

For a given mass of an ideal gas at constant pressure, therefore, the quantity V/T is constant; that is, the volume is proportional to the absolute temperature. This relationship, which is an alternative statement of Gay-Lussac's law, is the basis of an ideal gas scale of temperature, the degree being the same as that on the Centigrade scale, namely, one hundredth part of the difference in temperature between melting ice and boiling water at a pressure of one atmosphere.*

Instead of defining an ideal gas as one obeying the laws of Boyle and Gay-Lussac, it may be described as one to which Boyle's law is applicable, and whose internal energy is independent of its volume at all temperatures; it can be shown (p. 236) that these two postulates include Gay-Lussac's law. By combining the two relationships $PV = \text{constant}$ at definite temperature, and $V/T = \text{constant}$ for a given pressure, it follows that for a given mass of an ideal gas

$$PV/T = \text{constant}. \quad (26)$$

This relationship is frequently employed, in one form or another, for the conversion of the volume of a gas to any required temperature and pressure.

If the volume of a gas is maintained constant and its temperature raised, there will be an increase of pressure; if α_p is the coefficient of

* It may be noted that the zero on the Centigrade scale, i.e., 0°C. , is the freezing point of water in equilibrium, and hence saturated, with air at 1 atm.; the freezing point of pure, air-free water is then $+0.0023^\circ \text{C.}$ at 1 atm. pressure.

pressure increase, then by analogy with (22) it is possible to write

$$P_t = P_0(1 + \alpha_p t), \quad (27)$$

where P_t is the pressure at t° c. and P_0 at 0° c. From (26), P/T must be constant at constant volume, so that (27) becomes

$$T_t = T_0(1 + \alpha_p t), \quad (28)$$

and from the definition of absolute temperature T_t , as $273.16 + t$ or $1/\alpha_v + t$, it follows at once that α_p , the pressure coefficient, is identical with the volume coefficient α_v for an ideal gas.

Equation of State for an Ideal Gas.—The assumption of Avogadro's law, that equal volumes of all gases, under the same conditions of temperature and pressure, contain equal numbers of molecules, leads to the conclusion that the constant in (26) should be independent of the nature of the gas, provided 1 mole is considered. The volume occupied by this quantity of an ideal gas at a given temperature and pressure will always be the same, so that for 1 mole PV/T must be a universal constant, represented by the symbol R ; hence for 1 mole of a gas, irrespective of its nature, provided it behaves ideally,

$$PV = RT, \quad (29)$$

where R is known as the molar **gas constant**. This equation embodies the three laws applicable to an ideal gas, and it is consequently called the **ideal gas equation**.

The Gas Constant.—Since actual gases do not behave ideally, the evaluation of the gas constant involves either making an allowance for the deviations in real gases or extrapolation to very low pressures. By such methods (p. 309) it has been estimated that at 0° c., i.e., 273.16° Abs., and at a pressure of 1 atm., i.e., 76 cm. of mercury, a mole of an ideal gas would occupy 22.414 liters. These data are sufficient for the calculation of R , but before doing so it is important to consider its dimensions, and the units to be used to express it. Pressure, that is force/area, has the dimensions $ml^{-1}t^{-2}$, and volume is represented by l^3 , so that the dimensions of PV are ml^2t^{-2} , which are those of energy (p. 181). The gas constant, equal to PV/T , has thus to be expressed in units of energy per degree per mole; in physical chemistry the chief energy units employed for the purpose are (a) liter-atmospheres, (b) ergs, and (c) calories.

The required result in liter-atm. is obtained by expressing the volume of 1 mole of ideal gas in liters, viz., 22.414, when the pressure is 1 atm. and the temperature is 273.16° Abs.; hence,

$$R = \frac{1 \times 22.414}{273.16} = 0.08205 \text{ liter-atm. deg.}^{-1} \text{ mole}^{-1}.$$

For the value of R in ergs it is necessary to express the pressure in c.g.s. units, so that 1 atm., i.e., 76 cm. of mercury, becomes 76.0×980.66

$\times 13.595$, i.e., 1.0132×10^6 , dynes per sq. cm.,* and the volume of 1 mole of ideal gas is 22414 ml., or 22414.6 cc.; hence,

$$R = \frac{1.0132 \times 10^6 \times 22414.6}{273.16} = 8.314 \times 10^7 \text{ ergs deg.}^{-1} \text{ mole}^{-1}.$$

Since it has been already shown (p. 184) that 1 (defined) calorie is equivalent to 4.184×10^7 ergs, it follows that

$$R = \frac{8.314 \times 10^7}{4.184 \times 10^7} = 1.987 \text{ cal. deg.}^{-1} \text{ mole}^{-1}.$$

It is seen that the gas constant is very close to 2 cal. deg.⁻¹ mole⁻¹; this value is frequently employed in approximate calculations.

Difference in Heat Capacities.—By subtraction of (14) from (16),

$$C_P - C_V = \left(\frac{\partial H}{\partial T} \right)_P - \left(\frac{\partial E}{\partial T} \right)_V,$$

and since H is equal to $E + PV$, by definition, this equation can be written as

$$C_P - C_V = \left(\frac{\partial E}{\partial T} \right)_P + P \left(\frac{\partial V}{\partial T} \right)_P - \left(\frac{\partial E}{\partial T} \right)_V. \quad (30)$$

Further, according to (17), since dE is a complete differential,

$$dE = \left(\frac{\partial E}{\partial V} \right)_T dV + \left(\frac{\partial E}{\partial T} \right)_V dT.$$

Upon dividing by dT and applying the restriction of constant pressure,† the result is

$$\left(\frac{\partial E}{\partial T} \right)_P = \left(\frac{\partial E}{\partial V} \right)_T \left(\frac{\partial V}{\partial T} \right)_P + \left(\frac{\partial E}{\partial T} \right)_V. \quad (31)$$

Addition of (31) to (30) then gives

$$C_P - C_V = \left[P + \left(\frac{\partial E}{\partial V} \right)_T \right] \left(\frac{\partial V}{\partial T} \right)_P. \quad (32)$$

This is the general equation ‡ for the difference in the heat capacities of *any* system; for an ideal gas, however, it may be simplified because $(\partial E/\partial V)_T$ is zero (p. 191), so that

$$C_P - C_V = P \left(\frac{\partial V}{\partial T} \right)_P. \quad (33)$$

The quantity $(\partial V/\partial T)_P$ represents the increase in volume for 1° rise of temperature at constant pressure, so that the right hand side of (33) is equal to the work done by the gas in expansion when the temperature is raised by this amount. The view previously expressed, and tacitly assumed by J. R. Mayer

* The value of g , the acceleration due to gravity, is assumed to be that at sea level and a latitude of 45°; this is 980.66 cm. per sec. per sec.

† This procedure may be open to objection on strict mathematical grounds, but the same result can be obtained by a more rigorous method.

‡ For other forms of this general equation, see pp. 235 and 236.

in his calculation of the mechanical equivalent of heat, that the difference in the specific heats of a gas at constant pressure and volume is exactly equivalent to the mechanical work done in expansion, is only true for an ideal gas, that is when $(\partial E/\partial V)_T$ is zero. Nevertheless, provided the temperature and pressure of a real gas are not too near the conditions under which liquefaction occurs, the variation of internal energy with volume is relatively small and (33) may be assumed to apply.

If 1 mole of an ideal gas is under consideration, so that $PV = RT$, it follows on differentiation with respect to temperature, at constant pressure, that

$$P \left(\frac{\partial V}{\partial T} \right)_P = R \quad (34)$$

and so (33) gives

$$C_P - C_V = R, \quad (35)$$

where C_P and C_V now refer to the *molar* heat capacities of the gas. For an ideal gas, and approximately for a real gas, the difference between the molar heat capacities at constant pressure and constant volume should thus be 1.987 cal. deg.⁻¹.

Reversible Processes.—A particular method of carrying out thermodynamic changes is of considerable importance; the process is carried out infinitesimally slowly, so that the system is always in temperature and pressure equilibrium with its surroundings. For this reason, a reversible process has been defined as a succession of equilibrium states. A very minute change, in one direction or the other, of the factors, e.g., temperature and pressure, determining the equilibrium will cause the system to change in one direction or in the opposite direction. A process performed in this manner is said to be **thermodynamically reversible**. In general, a reversible process may be defined as *one which can be performed in opposite directions, so that all changes occurring in any part of the direct process are exactly reversed in the corresponding parts of the reverse process, and only infinitesimal changes are produced in external systems*. As a simple illustration of a reversible process, imagine a liquid in equilibrium with its vapor in a vessel closed by a piston and placed in a constant temperature bath. The pressure on the piston will then be equal to the vapor pressure of the liquid at the given temperature (cf. p. 443). If the external pressure is increased by an infinitesimally small amount the vapor will condense, but this process will take place at such a slow rate that the temperature of the system will not change, and the pressure above the liquid remains unaltered. Although condensation is occurring, therefore, the system at every instant is in equilibrium. Similarly, if the external pressure is made just smaller than the vapor pressure, the liquid will evaporate at an extremely slow rate, and although the system is changing, since evaporation is occurring, the change may be regarded as a series of equilibrium states. Rapid evaporation or condensation, resulting from sudden expansion or compression, will lead to temperature

changes and to a considerable disturbance of equilibrium and the process will not be reversible.

The work done in reversible vaporization or condensation at constant temperature can be readily calculated. In these processes the pressure remains constant throughout, since it is always equal to the equilibrium vapor pressure p of the liquid at the experimental temperature. If ΔV is the increase of volume in the course of the reversible evaporation, the work of expansion is equal to $p\Delta V$, done by the system [cf. equation (1)]. For the corresponding condensation the same amount of work would be done on the system. Since p and ΔV are both quite definite in these reversible processes, the work done is also definite. As there is no other form of work, it follows that the heat change in reversible vaporization and condensation must also have a definite value; this is in fact equal to the heat of vaporization (see p. 444) at the specified temperature.

Reversible Isothermal Expansion of Ideal Gas.—In the reversible evaporation or condensation process the pressure of the vapor remained constant, because it was always in equilibrium with the liquid. When a gas expands isothermally, however, the pressure must inevitably decrease in the course of the expansion. In order to carry out such a process in a reversible manner, it is supposed that the gas is enclosed in a cylinder fitted with a weightless and frictionless piston, and the external pressure on the latter is adjusted so as to be always less than that of the gas by an infinitesimally small amount. The expansion would then take place extremely slowly. If the cylinder is imagined to be placed in a large thermostat, the heat required by the gas to compensate for the work done in expansion would be available; since the process is carried out very slowly, the absorption of heat keeps pace with its utilization in the form of work, and the temperature of the system remains constant. If at any instant during the expansion the external pressure were increased, so that it became greater than the gas pressure by an extremely small amount, the process would be reversed; compression would now take place, at constant temperature, and every stage would be exactly the reverse of that which occurred during expansion. At the same time, there would be no more than infinitesimal changes in external bodies, as required by the definition given above.

If expansion had been carried out rapidly, the temperature of the gas would have fallen appreciably, and as a result there would have been a transfer of heat from the warmer reservoir to the cooler gas. When the process is reversed, that is, in compression, it would be necessary, therefore, for the heat to pass of its own accord from the colder to the warmer body; as will be seen later such a transfer is contrary to all experience. A rapid expansion or compression is thus an irreversible process in the thermodynamic sense.

The work done in an *isothermal*, reversible expansion is the maximum work possible for the given volume change; it can be calculated in the following manner. If P is the pressure of the gas at any stage in the ex-

pansion, the external pressure will be $P - dP$, where dP is a very small quantity. The work w which is done when the gas increases its volume by an infinitesimally small amount dV , is then equal to the product of the external pressure and the volume change; thus,

$$w = (P - dP)dV.$$

Neglecting the very small product $dPdV$, it follows that

$$w = PdV. \quad (36)$$

The total work done in the expansion will be the sum of a continuous series of PdV terms, with the value of P decreasing steadily, and the volume increasing, from that in the initial state to that in the final state. The total work of expansion W may then be expressed by the integral

$$W = \int_{V_1}^{V_2} PdV, \quad (37)$$

where P is the gas pressure at any instant; V_1 is the initial volume of the gas and V_2 is its final volume.

It should be remembered that in the general case the integral of (37) has no definite value, but for an isothermal, reversible process P is determined by the constant temperature of the gas and the volume occupied at any instant. In these circumstances the path between initial and final states is defined, and so also is the work of expansion. If the relationship between the pressure and volume is known, the integral may be evaluated so as to give the work involved in an isothermal and reversible process. For example, if the system consists of 1 mole of an ideal gas, then $PV = RT$, and $P = RT/V$; since the temperature remains constant, (37) takes the form

$$W = RT \int_{V_1}^{V_2} \frac{dV}{V} = RT \ln \frac{V_2}{V_1}. \quad (38)$$

Since PV is constant, i.e., $P_1V_1 = P_2V_2$, this result may be expressed alternatively as

$$W = RT \ln \frac{P_1}{P_2}. \quad (39)$$

It should be understood that (38) and (39) apply both to expansion and compression, provided P_1 and V_1 always represent the initial, and P_2 and V_2 the final state of the gas. For an expansion $V_2 > V_1$ and w is positive, while for a compression $V_1 > V_2$ and w has a negative value. In any event, it is clear from the equations that the work done depends only on the initial and final volumes (or pressures), and the constant temperature; hence, it is independent of the path, provided it is isothermal and reversible.

Equations (38) and (39) give the work done when the volume of 1

mole of gas is changed reversibly under isothermal conditions, i.e., heat enters or leaves the system so as to maintain the temperature T constant. For an ideal gas, the energy is independent of the volume at constant temperature, so that ΔE is zero; it follows, therefore, from the first law equation (5), that in an isothermal change involving an ideal gas

$$Q = W; \quad (40)$$

that is, the heat absorbed from its surroundings is equal to the work done upon them by the gas. If the gas is compressed then W is negative, and work is done *on* the gas; Q is then also negative and represents heat *evolved*, which is equal in amount to the work done.

It is of interest to call attention to the fact that in an isothermal change, the work of expansion is a maximum when the process is carried out reversibly. The only way in which the work could possibly be increased would be by increasing the external pressure. Since the latter is only infinitesimally less than the gas pressure, it is evident that if the external pressure is increased the expansion process would become impossible. Hence, the value of the work given by (38) or (39) is the maximum that could be obtained for the specified increase of volume at constant temperature. In the case of an isothermal compression, the reversible work is the maximum work done *on* the system, but the minimum done *by* the system.

Adiabatic Changes.—A special type of process which can be performed reversibly is that known as an **adiabatic change** (Greek: *not passing through*), in which it is supposed that no heat enters or leaves the system. In an infinitesimally slow adiabatic expansion or compression of a gas, for example, the temperature and pressure will adjust themselves automatically to the proper equilibrium values as the volume is changed; the process is then thermodynamically reversible. (The essential difference between isothermal and adiabatic changes is that in the former case the temperature remains constant and the system exchanges heat with its surroundings, while in the latter case there must be no heat exchange and so the temperature will be altered.) As explained below, it will fall during an adiabatic expansion and rise in a compression.

It is perhaps opportune to mention here that, although it may be possible in practice to approach reversibility, it is impossible to attain complete reversibility. As it is not feasible to construct a vessel which is absolutely impermeable to heat, so a perfect adiabatic change is impracticable. Similarly, an isothermal expansion or contraction cannot be performed *infinitesimally* slowly; in actual fact all such processes must occur at a finite, even if small, rate and there is a consequent departure from completely reversible behavior. Nevertheless, in spite of the impracticability of reversible changes the concept is of great theoretical importance.

For a small stage in a reversible adiabatic process q is zero, so that (6) takes the form

$$dE + w = 0. \quad (41)$$

If the work done is external work due to a volume change only, w may be replaced by PdV , so that

$$dE + PdV = 0. \quad (42)$$

Since $(\partial E/\partial V)_T$ is zero for an ideal gas, (17) becomes

$$dE = \left(\frac{\partial E}{\partial T} \right)_V dT = C_V dT, \quad (43)$$

because $(\partial E/\partial T)_V$ is equal to C_V ; it follows, therefore, from (42) and (43) that

$$C_V dT + PdV = 0 \quad (44)$$

for an adiabatic process with an ideal gas. In an adiabatic expansion dV is positive and hence dT must be negative, that is to say, the temperature falls; on the other hand, as a consequence of the adiabatic compression of a gas, its temperature will rise. These results are to be expected on general grounds: a gas does work in expansion and as no heat enters or leaves the system the energy is obtained at the expense of the internal energy of the gas, and so its temperature falls. Similarly, in a compression work is done on the gas and there will be a rise of temperature because of the increase of internal energy.

From (44) it is possible to deduce the quantitative relationship between the pressure, or temperature, and volume during an adiabatic change in the following manner. If 1 mole of ideal gas is considered, $P = RT/V$, so that (44) becomes

$$C_V dT + RT \frac{dV}{V} = 0 \quad (45)$$

or

$$C_V dT/T + R dV/V = 0, \quad (46)$$

where C_V is the molar heat capacity at constant volume. If P_1 , V_1 and T_1 are the pressure, volume and temperature at the commencement, and P_2 , V_2 and T_2 the values at the conclusion of the adiabatic process, then it follows by integration of (46), T changing from T_1 to T_2 while V changes from V_1 to V_2 , C_V being constant, that

$$C_V \ln T_2/T_1 + R \ln V_2/V_1 = 0. \quad (47)$$

Since 1 mole of gas is being considered, R may be replaced by $C_P - C_V$, and writing γ for C_P/C_V , the ratio of the two molar heat capacities, (47) can be reduced to

$$-\ln T_2/T_1 = (\gamma - 1) \ln V_2/V_1; \quad (48)$$

$$\therefore T_1/T_2 = (V_2/V_1)^{\gamma-1}. \quad (49)$$

As the gas is ideal, $P_1 V_1 = RT_1$ and $P_2 V_2 = RT_2$,

$$\therefore T_1/T_2 = P_1 V_1/P_2 V_2, \quad (50)$$

and from (49) and (50) it follows that

$$P_1 V_1^\gamma = P_2 V_2^\gamma; \quad (51)$$

hence for an adiabatic process PV^γ is constant, as compared with PV constant for an isothermal process. Since γ is greater than unity, it is evident that for a given decrease of pressure the increase of volume will be less in an adiabatic

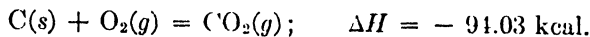
than in the corresponding isothermal expansion; that is to say, the plot of pressure against volume, i.e., the pressure-volume curve, will be steeper in the former than in the latter case (see p. 220). The reason for the smaller volume change in an adiabatic expansion is to be attributed to the accompanying fall in temperature; in an isothermal expansion the temperature remains constant.

THERMOCHEMISTRY

Heat Changes in Chemical Reactions.—The aspect of physical chemistry which deals with the heat changes accompanying chemical reactions is known as **thermochemistry** and, as will be seen shortly, this subject is based to a great extent on the law of conservation of energy. Since different substances have different amounts of energy, the total energy of the products of a chemical reaction may be different from the total energy of the reactants, so it is evident that the process may be accompanied by an absorption or liberation of energy in the form of heat. If heat is liberated in the reaction the process is said to be **exothermic**, but if heat is absorbed it is described as an **endothermic** reaction. The majority of, although not all, chemical reactions which go to virtual completion at ordinary temperatures are exothermic.

If a chemical reaction is accompanied by a volume change, as is particularly the case for many processes involving the combination of gases, the heat change will depend on whether the reaction is carried out at constant pressure or at constant volume. One or other of these conditions must be specified in order to make the heat absorbed or evolved a definite quantity which is independent of the reaction path. Since many reactions normally occur at constant (atmospheric) pressure, it is the usual practice to record heat changes in terms of Q_p , the heat absorbed at constant pressure; as seen above [equation (10)], this is, of course, equal to ΔH , the increase of heat content accompanying the process. This quantity is often referred to as the **heat of reaction**.* The value of Q_v (or ΔE), the heat absorbed at constant volume, can be readily derived from ΔH , as will be seen shortly.

The heat change accompanying a reaction, for example that between solid carbon (graphite) and gaseous oxygen to yield carbon dioxide gas, is represented in the form of a thermochemical equation; thus,

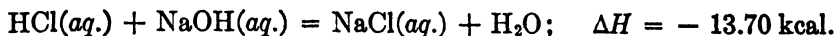


This means that when 12 g. of solid carbon and 32 g. of gaseous oxygen react to form 44 g. of gaseous carbon dioxide at constant pressure, there is a decrease in heat content of 94.03 kcal.,† or 94,030 cal.; in other words,

* In much of the older literature the heat of reaction is defined as the heat *evolved*, that is, $-\Delta H$.

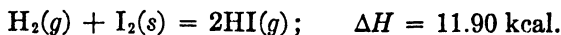
† Since heats of reaction are usually large, and can rarely be measured with an accuracy better than ± 100 cal., it is now the general practice to express thermochemical quantities in terms of kilogram calories, i.e., kcal, where 1 kcal. = 1000 cal.

this quantity of heat is evolved, since ΔH is negative. The symbol g , l or s , placed in parentheses after each formula, indicates whether the substance involved in the reaction was in the gaseous, liquid or solid state. Reactions taking place in aqueous solution are indicated by the use of the symbol $aq.$; thus,



Strictly speaking $aq.$ implies that the reaction is taking place in such dilute solution that the addition of further solvent causes no heat change. In other words, the conditions are such that the heat of dilution is zero; this is exactly true, however, only at infinite dilution.

The two reactions referred to above are exothermic, since in each case ΔH is negative and hence heat is evolved; an endothermic reaction, i.e., with ΔH positive, is the following:



Heat Changes at Constant Pressure and Constant Volume.—The relationship between the heat of reaction at constant pressure (ΔH) and at constant volume (ΔE) is given by (11), which may be written as

$$\Delta E = \Delta H - P\Delta V, \quad (52)$$

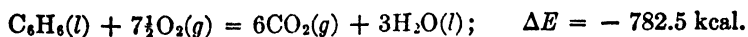
where ΔV is the increase of volume when the reaction occurs at constant pressure P . The quantity $P\Delta V$ represents the external work done by the system when the reaction is carried out at constant pressure; at constant volume no external work is done, and the heat absorbed (or evolved) is equal merely to the increase (or decrease) of the internal energy E . For a reaction involving solids and liquids only, the volume change ΔV is usually so small that $P\Delta V$ is negligible; ΔH and ΔE may then be taken as equal within the limits of experimental error. If gases are concerned, however, the change in volume may be quite large; in such cases the difference between ΔH and ΔE can be readily calculated in the following manner. If V is the volume of 1 mole of gas, and $\Delta \nu$ is the increase in the number of gas molecules in the reaction, then ΔV is equal to $V\Delta \nu$; hence (52) becomes

$$\Delta E = \Delta H - PV\Delta \nu. \quad (53)$$

Assuming the gases to behave ideally, PV may be replaced by RT , so that

$$\Delta E = \Delta H - RT\Delta \nu. \quad (54)$$

The application of these equations may be illustrated by reference to the combustion of liquid benzene; measured at constant volume and 300° Abs. ,



In this reaction $\Delta \nu$ is -1.5 , since there are 7.5 moles of gaseous reactants and 6 moles of gaseous products; taking R as 2 cal., i.e., 2×10^{-3} kcal., it is seen

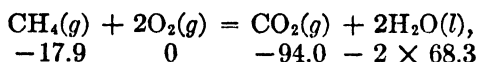
that

$$\begin{aligned}\Delta H &= \Delta E + RT\Delta v \\ &= -782.5 - (2 \times 10^{-3} \times 300 \times 1.5) = -783.4 \text{ kcal.}\end{aligned}$$

The heat of reaction at constant pressure, at 300° Abs., is thus -783.4 kcal.

Heat of Formation.—The heat of formation of a compound is the increase of heat content ΔH when 1 mole of the substance is formed from its elements. It is usually postulated that the elements are in their so-called **standard states**, that is, in their stable forms at ordinary temperature and 1 atm. pressure. From the examples given above it is evident that the heat of formation of carbon dioxide is -94.03 kcal.; since graphite is the stable form of carbon at normal temperatures and pressures, it represents the standard state of carbon. Similarly, the heat of formation of hydrogen iodide is $\frac{1}{2} \times 11.90$, i.e., 5.95 kcal., since solid iodine is the standard state of this element.

The heats of formation of the compounds involved in a reaction have an important connection with the heat of reaction. For purposes of calculation, the convention is adopted that the heat contents of all elements in their standard states are arbitrarily taken to be zero at all temperatures. The heat content of a compound is then equal to its heat of formation ΔH . By means of this convention, and the corollary just given, it is possible to calculate the heat change in any reaction, provided the heats of formation of the substances taking part in it are known. Consider the reaction



the heat content, i.e., heat of formation, of each species being written below its formula. The change in heat content for the whole reaction is then given by

$$\Delta H = [-94.0 + (-2 \times 68.3)] - [(-17.9) + 0] = -212.7 \text{ kcal.}$$

Instead of using heats of formation to calculate a heat of reaction, the procedure may be reversed; the heat of formation of a compound taking part in a reaction can be calculated, provided the values for all the other compounds and the heat of the complete reaction are known. The heats of formation of a number of familiar compounds are recorded in Table 26.

TABLE 26. HEATS OF FORMATION PER MOLE IN KCAL. AT 25° C.

Substance	ΔH	Substance	ΔH	Substance	ΔH
HBr	-8.4	N ₂ O	17.0	C ₂ H ₆	-23.4
HI	6.0	NH ₃	-11.0	CH ₃ OH(l)	-60.0
SO ₂	-70.2	CO	-26.4	C ₂ H ₅ OH(l)	-66.2
H ₂ S	-5.2	C ₂ H ₂	54.3	C ₂ H ₄ O ₂ (l)	-118.0
NO	21.5	C ₂ H ₄	9.6	HCN(g)	30.2

Heats of reaction, and hence heats of formation, vary with temperature to some extent (cf. p. 212). As far as possible efforts are now made

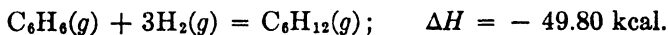
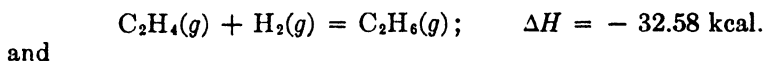
to record data applicable to the uniform temperature of 25° C., i.e., 298° Abs.; this is often indicated by inserting the absolute temperature as a subscript, e.g., ΔH_{298° .

Heat of Combustion.—Organic compounds containing carbon, hydrogen and oxygen only can be burnt in oxygen to give carbon dioxide and water as the sole products; the heat content change accompanying the complete combustion of 1 mole of a compound is called its **heat of combustion**.* From data given earlier, the heats of combustion of benzene are -783.4 kcal., and that of methane is -212.7 kcal., respectively, the values for other compounds are quoted in Table 27. It will be shown later that heats of combustion are of importance in connection with the calculation of heats of formation of organic substances when direct calorimetric measurement is not possible.

TABLE 27. HEATS OF COMBUSTION PER MOLE IN KCAL. AT 25° C.

Substance	$-\Delta H$	Substance	$-\Delta H$
Benzene(l)	783.4	Methyl alcohol(l)	170.9
Ethane(g)	368.4	Ethyl alcohol(l)	327
Cyclohexane(l)	939	Acetic acid(l)	209
n-Hexane(l)	990.6	Diethyl ether(l)	651.7
Toluene(l)	936	Acetone(l)	427
Sucrose(s)	1,360	Ethyl acetate(l)	538

Heat of Hydrogenation.—In recent years, mainly as a consequence of the work of G. B. Kistiakowsky and his collaborators (1935 *et seq.*), the **heat of hydrogenation** of unsaturated compounds has attracted attention (see p. 591). It may be defined as the increase in heat content when 1 mole of an unsaturated hydrocarbon is converted into the corresponding saturated compound by gaseous hydrogen; the following examples are for the hydrogenation of ethylene and benzene.



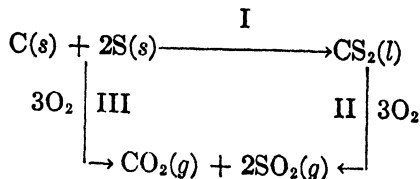
Thermochemical Laws.—The calculations of heats of reaction and of formation given above have been tacitly based on the principle of the conservation of energy, and hence on the first law of thermodynamics. Two important thermochemical laws are based on the same principle. According to the observations of A. L. Lavoisier and P. S. de Laplace (1780), who made the earliest quantitative studies in thermochemistry, *the quantity of heat required to decompose a compound into its elements is equal to the heat evolved when that compound is formed from its elements*; that is, the **heat of decomposition** of a compound is numerically equal to its heat of formation, but of opposite sign. This is the earliest recorded thermochemical generalization; it is, of course, to be expected from the

* Some heats of combustion of organic compounds containing nitrogen and halogens have been measured, but the results are of doubtful significance because of the uncertainty as to the nature of the products.

first law of thermodynamics, for otherwise it would be possible to create or destroy heat energy by making a compound from its elements and then decomposing it, or *vice versa*. Another aspect of the same general principle is that thermochemical equations may be reversed, provided the sign of the heat term is changed, its numerical value remaining the same.

The second important law of thermochemistry was discovered experimentally by G. H. Hess in 1840; it states that *the heat change in a chemical reaction is the same whether it takes place in one or several stages*. This means that the net heat of reaction depends only on the initial and final states, and not on the intermediate states through which the system may pass. As was first realized by J. Thomsen (1853), the law of Hess, often called the **law of constant heat summation**, is a direct consequence of the law of conservation of energy. It has been shown above that ΔH and ΔE are dependent only on the initial and final states (p. 188), and these are equal to the heats of reaction at constant pressure and constant volume respectively. Let *A* represent the reactants in a process and *B* the products, and suppose it is possible to go from *A* to *B* by two alternative paths; if the heat changes in these two paths are not the same, then by going from *A* to *B* by one path and returning to the original state by the other, it would be possible to create energy. Since this is considered to be impossible, according to the first law, it follows that the heat produced in the reaction must be independent of the method whereby the change is carried out, provided either the pressure or the volume is maintained constant.

An important consequence of Hess's law is that thermochemical equations may be added and subtracted, with the result that heats of reaction which are not directly accessible can be calculated. It is often required to determine the heat of formation of an organic compound from its elements although it cannot be measured directly; if the heats of combustion of the compound and of its constituent elements are known, however, the application of the law of constant heat summation allows the requisite information to be obtained. Starting with the elements carbon and sulfur, it is possible to combine them to form carbon disulfide which can be burnt in oxygen to give carbon dioxide and sulfur dioxide, or the carbon and sulfur can be converted directly into the corresponding oxides; thus,

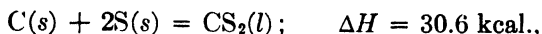


If ΔH_{I} , ΔH_{II} and ΔH_{III} are the heat content changes in the respective processes marked I, II and III, then it follows from Hess's law that $\Delta H_{\text{I}} + \Delta H_{\text{II}} = \Delta H_{\text{III}}$; the same net heat change being involved in the

direct oxidation as in the indirect one through carbon disulfide. The thermochemical quantities are as follows:

- (i) $\text{CS}_2(l) + 3\text{O}_2(g) = \text{CO}_2(g) + 2\text{SO}_2(g); \quad \Delta H = -265.0 \text{ kcal.}$
 (ii) $\text{C}(s) + \text{O}_2(g) = \text{CO}_2(g); \quad \Delta H = -94.0 \text{ kcal.}$
 (iii) $2\text{S}(s) + 2\text{O}_2(g) = 2\text{SO}_2(g); \quad \Delta H = -2 \times 70.2 \text{ kcal.}$

The value of ΔH_{II} is thus -265.0 , and ΔH_{III} is $-(94.0 + 2 \times 70.2)$, i.e., -234.4 kcal. ; since $\Delta H_{\text{I}} = \Delta H_{\text{III}} - \Delta H_{\text{II}}$, it follows that ΔH_{I} is $-234.4 - (-265.0) = 30.6$, and the heat of formation of carbon disulfide from atomic carbon and sulfur is 30.6 kcal. The same conclusion may be reached in an alternative manner by adding the thermochemical equations (ii) and (iii) and subtracting (i); the result is

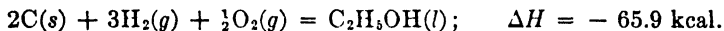


as before.

The heats of formation of hydrocarbons and of compounds of carbon, hydrogen and oxygen, may be evaluated in an analogous manner; the data for ethyl alcohol may be taken in order to illustrate the method.

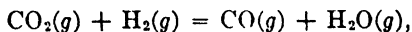
- (i) $\text{C}_2\text{H}_5\text{OH}(l) + 3\text{O}_2(g) = 2\text{CO}_2(g) + 3\text{H}_2\text{O}(l); \quad \Delta H = -327.0 \text{ kcal.}$
 (ii) $\text{C}(s) + \text{O}_2(g) = \text{CO}_2(g); \quad \Delta H = -94.0 \text{ kcal.}$
 (iii) $\text{H}_2(g) + \frac{1}{2}\text{O}_2(g) = \text{H}_2\text{O}(l); \quad \Delta H = -68.3 \text{ kcal.}$

If (ii) is multiplied by two and (iii) multiplied by three is added, and then (i) subtracted, it is seen that



The heat of formation of 1 mole of ethyl alcohol from its elements is thus -65.9 kcal.

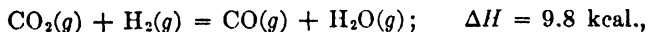
Not only heats of formation, but the heat changes of other reactions can be calculated by the use of Hess's law; for example, if it is required to find the change of heat content for the process



three separate changes, all of which can be studied directly, may be considered.

- (i) $\text{C}(o) + \frac{1}{2}\text{O}_2(g) = \text{CO}(g); \quad \Delta H = -67.6 \text{ kcal.}$
 (ii) $\text{H}_2(g) + \frac{1}{2}\text{O}_2(g) = \text{H}_2\text{O}(l); \quad \Delta H = -68.3 \text{ kcal.}$
 (iii) * $\text{H}_2\text{O}(g) = \text{H}_2\text{O}(l); \quad \Delta H = -10.5 \text{ kcal.}$

If (i) is reversed and then (ii) added and (iii) subtracted, it follows that



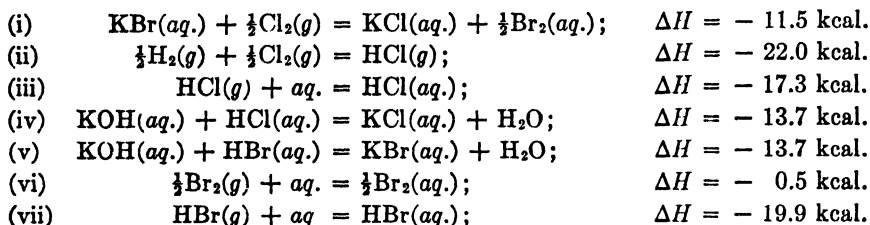
so that the heat absorbed in the reaction, calculated for ordinary temperatures, is 9.8 kcal.

In general, heats of combustion may be used to evaluate heats of reaction. If the sum of the heats of combustion of the products is subtracted from the

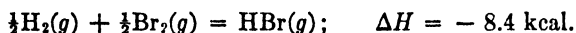
* This process is the condensation of steam; ΔH is thus numerically equal to the heat of vaporization per mole of water at ordinary temperatures.

heat of combustion of the reactants, the difference gives the heat evolved in the process. The reaction between carbon dioxide and hydrogen just examined and the examples of the application of Hess's law considered above will be seen to be illustrations of this general rule.

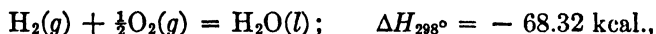
In conclusion, a case of an entirely different type showing the use of Hess's law will be given; it is taken from the classical work of J. Thomsen, and shows how the heat of formation of gaseous hydrogen bromide was evaluated. The thermal changes in the following reactions were measured, with the results quoted; it will be noted that (iii), (vi) and (vii) represent heats of solution.



If the thermochemical equations (ii), (iii), (iv) and (vi) are added, and from the total the sum of (i), (v) and (vii) is subtracted, the result is



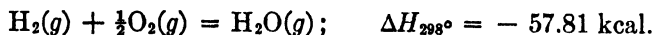
Influence of Physical State.—The heat change in a reaction obviously depends on the physical state, e.g., solid, liquid or gaseous, of the substances involved; a very simple example occurs in the formation of water. If the latter is in the liquid state, then the reaction is given by



but if water vapor is the product then the heat of reaction is obtained by subtracting the molar heat of vaporization; thus,

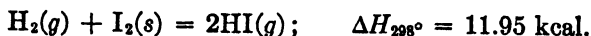


and so,

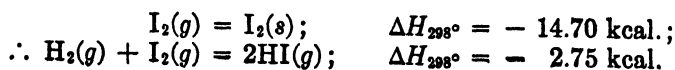


The heat of formation of 1 mole of water vapor at 298° Abs. , i.e., 25° C. , is thus -57.81 kcal.

When the heat of a reaction involving solid substances is known and the value for the liquid state is required, the necessary allowance can be made by means of the latent heat of fusion; similarly, allowance for the difference between the heat changes for solid state and vapor is made by utilizing the heat of sublimation (see p. 459). As already seen,



and the heat of sublimation of iodine at ordinary temperatures is given by



Different heat changes are involved for different allotropic modifications of an element, or for different crystalline forms of a compound; if the data for a given reaction involving the two forms are available, the heat of transition of one form to the other may be evaluated. For example, the heats of combustion of the two allotropic forms of carbon, viz., diamond and graphite, are -94.48 and -94.03 kcal., respectively; it follows, therefore, that

$$C(\text{diamond}) = C(\text{graphite}); \quad \Delta H = -0.45 \text{ kcal.}$$

The heat evolved when the so-called "amorphous" carbon, charcoal, is burnt is about 97.0 kcal., although this substance has probably the same fundamental crystal structure as graphite; there is thus a considerable difference of heat content as the result of a change in the size of the crystallites. The finely divided form, charcoal, has a higher heat content than the one consisting of larger aggregates.

Experimental Methods: Types of Calorimeters.—The heat changes involved in chemical reactions are measured by carrying out the process in a suitable manner in a calorimeter; in order to obtain reasonably accurate results, however, it is necessary to take many precautions. The chief points requiring attention are efficient stirring so as to ensure uniformity of temperature in the calorimetric liquid, avoidance of evaporation of the latter, reduction to a minimum of convection currents in the air, and correction for, or elimination of, heat losses by radiation. Two main types of calorimeter have been in use; in the first, the aim is to keep the liquid in the surrounding vessel at a constant temperature, whereas in the second the temperature of the outer jacket is raised as required, so that it is always the same as that of the liquid in the calorimeter. In his important work on thermochemistry, J. Thomsen used the first type; his actual apparatus varied with the nature of the reaction being studied, but the calorimeter generally consisted of a platinum vessel, having a capacity of a liter or more, surrounded by another platinum vessel with an air gap between. The whole was kept in a room the temperature of which varied by no more than 0.1° or 0.2° in the course of a day. M. Berthelot, on the other hand, surrounded his platinum calorimeter proper, having a capacity of a half to one liter, with a thin copper vessel, and this was again surrounded by one of silver; these were then placed in a large double-walled container, the space between the walls being filled with water, and the whole surrounded by a thick layer of felt to minimize losses by radiation.

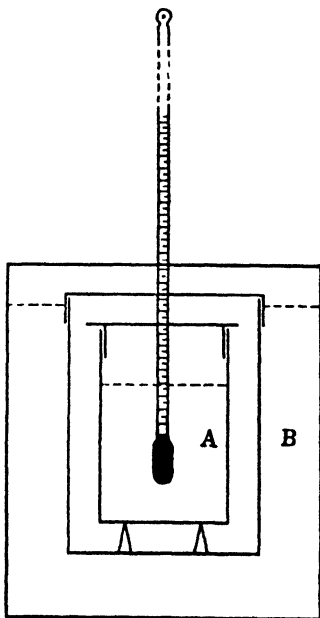


FIG. 35. Simple form of calorimeter

An apparatus of this form, with perhaps omission of one of the intermediate vessels, has been often employed in thermochemical work. According to W. P. White (1928), who has made a thorough analysis of the errors involved in calorimetry, very good results may be obtained with the apparatus shown diagrammatically in Fig. 35. The metal vessel *A* is the calorimeter, and this is separated by an air gap of about 1 cm. from the outer jacket *B* containing water kept at a constant temperature by a thermostatic device. The use of a wider air gap would decrease radiation losses, but increase errors due to convection currents. The liquids in both *A* and *B* are kept well mixed by means of stirrers, not shown in the diagram. Evaporation and convection are minimized by the use of metal covers, and radiation effects are diminished by surrounding the whole apparatus with insulating material. Results of a fair order of accuracy have been obtained by using a Dewar vessel as a calorimeter; heat losses due to radiation are small, and can be made even less by placing the vessel in a box coated with felt.

With the type of apparatus described there are always heat losses by radiation, even though they may be small, from the calorimeter to the surroundings; the object of keeping the temperature of the outer jacket constant is to maintain a steady rate of heat loss. In order to make the necessary allowance for this, the temperature of the calorimetric liquid is read at regular intervals for a little time before carrying out the reaction, and also after the reactants have been brought together; by extrapolating the two sets of data to the moment of mixing, the former gives the initial and the latter the corrected final temperature of the calorimeter, the difference being the true rise of temperature. In special cases, however, a more accurate procedure must be used (W. P. White, 1926).

The water equivalent (heat capacity) of the calorimeter may be evaluated by multiplying the weights of the various parts of the calorimetric vessel, thermometer, stirrer, etc., by their respective specific heats; to this is added the weight of water in the vessel, and the sum is multiplied by the rise of temperature to give the heat evolution in the reaction. In modern work, the heat capacity of the calorimeter is almost invariably determined by electrical heating. A coil of known resistance is placed in the calorimeter, and a known current of electricity is passed through it for a measured time; the resulting rise of temperature is then observed. The quantity of heat liberated is RI^2t int. joules, where R is the resistance of the coil in int. ohms, I is the current in int. amps., and t is the time in sec. for which it is passed. If this is divided by the rise of temperature, the result is the heat capacity of the calorimeter and its contents in int. joules per degree. The measurement of the heat capacity should be made under conditions that approximate as closely as possible to those existing in the thermochemical experiment.

The Adiabatic Calorimeter.—The second type of calorimeter to which earlier reference was made was proposed by T. W. Richards (1905) and perfected by him and his co-workers; it is known as the **adiabatic calorimeter**, and has been employed in one form or another in most of the accurate work of recent times. Since the temperature of the outer jacket is kept as close as possible to, that is, not more than 0.1° different from, that of the calorimeter vessel the heat losses due to radiation and convection are negligible, and no correction is necessary. The greatest advantage of the adiabatic method is realized when relatively slow reactions are being studied; for such processes the correction for radiation losses,

as described above, are unreliable. In the first adiabatic calorimeter the outer jacket contained dilute alkali and its temperature was raised in a gradual manner by the slow addition of acid, as required, but in later models electrical heating was used. The best way to obtain uniform heating is to pass the current electrolytically through the liquid in the jacket; the metal calorimeter and the outer jacket act as electrodes. With the aid of a thermocouple, one junction of which is in the calorimetric liquid and the other in the outer vessel, and suitable relays, the adjustment of the temperature can be made entirely automatic. In a modified form of the adiabatic calorimeter, D. A. MacInnes and J. M. Braham (1917) adjusted the temperature of the outer jacket so that it was always 0.15° below that of the inner vessel; the small amount of heat lost by radiation was then found to balance exactly the heat generated by the stirrer. It was thus possible to avoid the appropriate correction for this quantity which is normally necessary for accurate results with calorimeters of all types. The heat capacity of the adiabatic calorimeter can be determined by either of the methods already described.

Measurement of Temperature.—For the measurement of the temperature rise in the calorimeter, calibrated thermometers are frequently employed, but a greater accuracy is attainable with thermal junctions or resistance thermometers. Differential calorimeters have been devised for the measurement of small heat changes. The one employed by E. Lange (1930) for the determination of heats of dilution of solutions is of particular interest, as it is probably capable of development for other purposes. A large Dewar vessel, divided axially into two equal water-tight compartments by a plate of hard rubber, acts as the calorimeter; it is immersed in a water bath, the temperature of which is adjusted according to the adiabatic principle. A 1000-element thermocouple is fixed in the dividing plate, a set of junctions being on each side, so that any difference of temperature between the two compartments can be immediately recorded. One section contains water, whereas the reaction occurs in the other. Very small changes of temperature are measured directly, but larger amounts of heat are determined by warming the comparison compartment with an electrical heating coil.

Types of Reaction.—For reactions in solution the process may be carried out directly in the calorimeter or in a thin-walled vessel immersed in it; the liquids should be mixed as rapidly as possible, the initial temperature being taken as the mean of the values for the two solutions. When gases or solids are involved a special technique has to be adopted. If the reaction between gases is relatively slow a catalyst may be employed, for it cannot affect the ultimate heat change; this procedure has been adopted in the work on the hydrogenation of unsaturated hydrocarbons (p. 203). In the determination of the heat of combination of hydrogen and chlorine, J. Thomsen passed the gases in a steady stream into a platinum bulb immersed in the calorimeter; the reaction was initiated by an electric spark and the hydrogen chloride was removed by an exit tube and its amount estimated. From this value, and the rise of temperature of the water in the calorimeter, the heat of formation of hydrogen chloride was determined, after making appropriate allowance for radiation losses. Heats of combustion of solids which burn readily, e.g., sulfur and hydrocarbons, were determined by M. Berthelot by ignition of the substance placed in a small crucible in a thin hard-glass vessel of about 400 cc. volume into which oxygen was passed. The products of burning were expelled through a glass

spiral tube surrounding the combustion vessel, so that the heat of reaction was completely transferred to the water in the calorimeter.

The Bomb Calorimeter.—The original arrangements of Berthelot, and of Thomsen, in which the substance was burnt in a special burner inserted into the bottom of the calorimeter, are now only of historical interest because of the invention by M. Berthelot (1881) of the **explosion bomb**. Although various slight modifications of this apparatus have been made, e.g., the Mahler, Kroeker, and Parr bombs, it remains substantially the same in principle as when first described. The organic substance is burned in oxygen under pressure, so that the reaction is rapid and complete; the heat developed is measured in a calorimeter in the usual manner. The Berthelot type of bomb is a cylindrical vessel, made of strong steel or other alloy, which can be closed with a pressure-tight cover; the capacity is about 400 or 500 cc., and the approximate external dimensions are height 15 cm., diameter 8 cm., and thickness of walls 0.6 cm. The bomb is lined internally with enamel, gold or platinum, to prevent oxidation of the steel. From two supports are suspended a platinum crucible into which is placed a weighed quantity of the substance to be burned; if the latter is volatile, it is enclosed in a thin glass tube. The supports also serve the purpose of making electrical connection, and they are joined together by a thin spiral of iron wire in contact with the material being studied. After screwing on the top, the bomb is filled with oxygen at about 25 atm. pressure or more, and the whole placed in a calorimeter. Fig. 36, taken from a publication by T. W. Richards (1907), shows the bomb in an adiabatic calorimeter; the stirring devices, electrical connections, and burette for the addition of acid to dilute alkali in the outer jacket, for the purpose of maintaining adiabatic conditions, can be seen.

When temperature equilibrium is attained, a current is passed through the iron wire which burns in the oxygen and so ignites the organic compound. Although combustion is not instantaneous, as is frequently supposed, it is generally very rapid, and the heat evolved is determined by the rise in temperature of the calorimetric liquid. A number of corrections have to be made, the chief being for the heat produced by the electric current and by the burning of the iron wire, for the formation of nitric acid because of the presence of nitrogen from the atmosphere or in the oxygen used, and for the heat developed by the stirring arrangement. In more recent work the iron wire device for igniting the compound has been replaced by a very thin platinum wire which fuses and sets fire to a thread of cotton; the correction is smaller than is the case when iron is employed. The heat capacity of the calorimeter and its contents is determined either by the electrical method already described, or better, according to the recommendations of the International Committee on Thermochemistry, by burning a substance of accurately known heat of combustion. The primary standard recommended for the purpose is benzoic acid, 1 g. of which is assumed to evolve 6319 cal. at 20° c. when burnt completely; * secondary standards, such as succinic acid, salicylic acid, naphthalene and sucrose, have been proposed. If for any reason it should be found necessary subsequently to alter the value for the primary standard, experimental data based on it can be readily adjusted.

Soon after its invention, the Berthelot bomb was used by F. C. A. Stoh-

* According to the U. S. Bureau of Standards (1942), the heat of combustion of benzoic acid is -3226.39 ± 0.32 int. kilojoules per mole at 25° c.

mann and his collaborators (1889 *et seq.*) for the determination of heats of combustion; more recently it has been employed by T. W. Richards, W. A. Roth, and by many others. It is of interest to note that bombs of similar type are in general use for the determination of the calorific values of solid and liquid fuels.

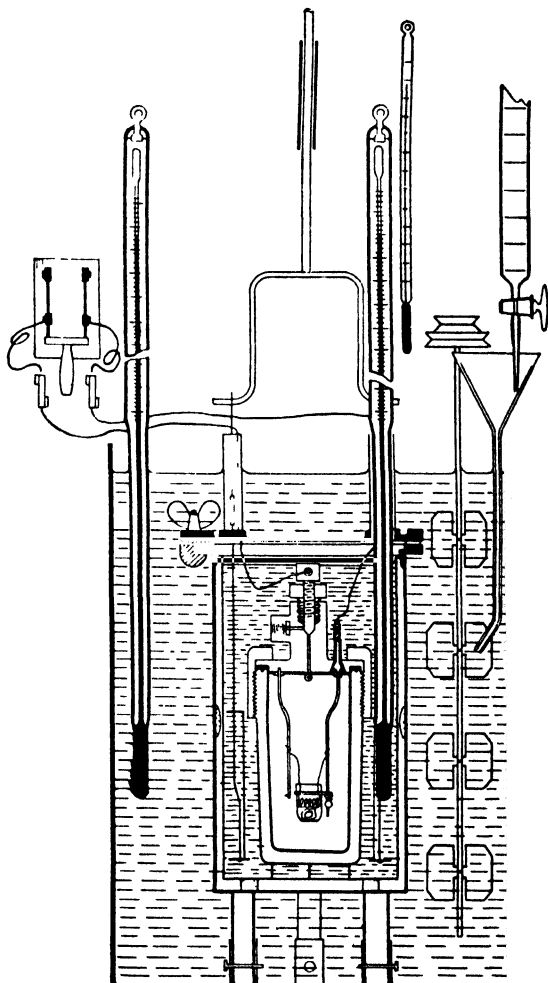


FIG. 36. Combustion bomb and calorimeter

The Thomsen-Berthelot Principle.—As a consequence of their extended studies in thermochemistry, J. Thomsen (1854) and M. Berthelot (1867) put forward the suggestion that the heat evolved in a chemical reaction was a measure of the “affinity” of the reacting substances; in other words, it was believed that only exothermic reactions could occur spontaneously. A little consideration shows that this view must be incorrect; for a chemically * re-

* Not to be confused with thermodynamic reversibility.

versible change the reactions must be able to occur both in the forward and reverse directions. If in one direction the process is exothermic, then it must be endothermic to the same extent in the opposite direction. It is clear, therefore, that in every chemically reversible system there is one reaction which is endothermic, yet it must take place spontaneously. Although the Thomsen-Berthelot principle is fundamentally incorrect, it is in many instances a close approximation, and it is true that at ordinary temperatures the reactions which take place most readily are those accompanied by an evolution of heat. It must be emphasized, however, that this is only a very rough generalization. Although exothermic reactions predominate at ordinary temperatures, endothermic changes become more common as the temperature is raised.

Kirchhoff's Equation: Influence of Temperature on Heat of Reaction.

—The heat change accompanying any process, physical or chemical, generally varies with temperature, and the relationship between this variation and other properties of the system concerned can be easily derived from the first law of thermodynamics. Consider any change represented by $A \rightarrow B$, where A is the initial state, i.e., the reactants, and B is the final state, i.e., the products. Suppose it is required to start with A at temperature T_1 and convert this into B at T_2 ; two methods are possible, but according to Hess's law the total heat change must be the same in both cases.

(1) The change $A \rightarrow B$ is made to occur at T_1 and constant pressure; the heat absorbed in the reaction is equal to the difference in the heat contents of the final and initial states at this temperature, i.e., $(H_B - H_A)_1 = \Delta H_1$. The temperature of the state B is now raised from T_1 to T_2 , that is, by ΔT , when the heat supplied will be $\Delta T(C_P)_B$, where $(C_P)_B$ is the heat capacity of B at constant pressure. The total heat change is thus

$$\Delta H_1 + \Delta T(C_P)_B.$$

(2) Alternatively, the temperature of A is first raised from T_1 to T_2 , and the heat absorbed is $\Delta T(C_P)_A$, where $(C_P)_A$ is the heat capacity of the state A ; the reaction is now made to take place at T_2 , when the increase of heat content $(H_B - H_A)_2 = \Delta H_2$ is equal to the heat absorbed. The total heat change at constant pressure is now

$$\Delta H_2 + \Delta T(C_P)_A.$$

According to the first law of thermodynamics the two heat changes must be identical, so that

$$\Delta H_1 + \Delta T(C_P)_B = \Delta H_2 + \Delta T(C_P)_A;$$

$$\therefore \frac{\Delta H_2 - \Delta H_1}{\Delta T} = (C_P)_B - (C_P)_A = \Delta C_P,$$

where ΔC_P is the difference in the heat capacities of final and initial states. If ΔT represents a small increase of temperature, this equation may be

written

$$\left(\frac{\partial(\Delta H)}{\partial T}\right)_P = \Delta C_P, \quad (55)$$

and since ΔH is equal to the heat of reaction at constant pressure, this relationship, first developed by G. R. Kirchhoff (1858), relates the variation of heat of reaction with temperature to the change in heat capacity accompanying the process.

The same result may be obtained in a simpler manner by utilizing the equations for heat capacity derived on p. 189. The increase in heat content ΔH in a process is equal to $H_B - H_A$, and differentiation with respect to temperature at constant pressure gives

$$\left(\frac{\partial(\Delta H)}{\partial T}\right)_P = \left(\frac{\partial H_B}{\partial T}\right)_P - \left(\frac{\partial H_A}{\partial T}\right)_P. \quad (56)$$

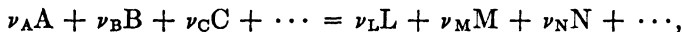
According to (16), $(\partial H/\partial T)_P$ is equal to C_P , and so (56) is seen to give (55) directly.

By means of arguments similar to either of those given above an analogous form of Kirchhoff's equation for a reaction at constant volume can be readily deduced. The heat absorbed is equal to the increase of internal energy, and the heat capacities involved are those at constant volume; hence,

$$\left(\frac{\partial(\Delta E)}{\partial T}\right)_V = \Delta C_V, \quad (57)$$

where ΔE is equal to the heat of reaction at constant volume, and ΔC_V is the increase of heat capacity at constant volume.

Although the Kirchhoff equation applies to both physical and chemical processes, its use will be examined here with special reference to chemical changes; consider, for example, the general reaction



where ν_A , ν_B , ν_C , etc. are the numbers of molecules of reactants A, B, C, etc., taking part, whereas ν_L , ν_M , ν_N , etc. represent the molecules of products L, M, N, etc. If ΔH is equal to the change of heat content for the quantities represented by the chemical equation, then

$$\Delta C_P = [\nu_L(C_P)_L + \nu_M(C_P)_M + \cdots] - [\nu_A(C_P)_A + \nu_B(C_P)_B + \cdots], \quad (58)$$

where C_P is the *molar* heat capacity of the species indicated by the subscript; the first term on the right hand side gives the heat capacity of all the molecules of products and the second term is that for all the reacting molecules, so that the difference is the increase in heat capacity for the reaction. These are the values of ΔH and ΔC_P to be used in (55).

If (55) is integrated between 0 and T , it is seen that

$$\Delta H - \Delta H_0 = \int_0^T \Delta C_P dT \quad (59)$$

or

$$\Delta H = \Delta H_0 + \int_0^T \Delta C_P dT, \quad (60)$$

ΔH being the increase in heat content at temperature T , and ΔH_0 , which is an integration constant, is equal to the hypothetical value at the absolute zero. In order to evaluate the integral in (60) it is necessary to know how the heat capacities of reactants and products vary with temperature. In general, an equation of the type

$$C_P = a + bT + cT^2 + dT^3 + \dots, \quad (61)$$

where a , b , c , etc., are constants for a given substance, is applicable. It follows, therefore, that

$$\begin{aligned} \Delta C_P &= \Sigma \nu a + (\Sigma \nu b)T + (\Sigma \nu c)T^2 + (\Sigma \nu d)T^3 + \dots \\ &= \alpha + \beta T + \gamma T^2 + \delta T^3 + \dots, \end{aligned} \quad (62)$$

where

$$\begin{aligned} \alpha &= \Sigma \nu a = (\nu_L a_L + \nu_M a_M + \dots) - (\nu_A a_A + \nu_B a_B + \dots), \\ \beta &= \Sigma \nu b = (\nu_L b_L + \nu_M b_M + \dots) - (\nu_A b_A + \nu_B b_B + \dots), \\ \gamma &= \Sigma \nu c = \text{etc.} \end{aligned}$$

The symbols a_A , b_B , etc., refer to the reactant A; a_B , b_B , etc., to the reactant B; etc., whereas a_L , b_L , etc., are the constants in the heat capacity equations for the product L, etc. If the value of ΔC_P as given by (62) is inserted in (60), it follows that

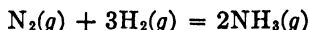
$$\Delta H = \Delta H_0 + \int_0^T (\alpha + \beta T + \gamma T^2 + \delta T^3 + \dots) dT \quad (63)$$

$$= \Delta H_0 + \alpha T + \frac{1}{2} \beta T^2 + \frac{1}{3} \gamma T^3 + \frac{1}{4} \delta T^4 + \dots, \quad (64)$$

which is the general equation for the variation of ΔH with temperature. It will be evident that if the heat capacities are independent of temperature, or may be assumed to be so over a small range of temperature, β , γ , δ , etc., are zero, and then

$$\Delta H = \Delta H_0 + \alpha T. \quad (65)$$

The application of the Kirchhoff equation may be illustrated by an actual example. For the reaction



the molar heat capacities, in the gaseous state, of the substances involved are

$$\begin{array}{ll} \text{N}_2 & C_P = 6.5 + 10^{-3} T \text{ cal. deg.}^{-1}, \\ \text{H}_2 & C_P = 6.5 + 9 \times 10^{-4} T, \\ \text{NH}_3 & C_P = 8.04 + 7 \times 10^{-4} T + 5.1 \times 10^{-6} T^2. \end{array}$$

According to (58), using these values for the heat capacities,

$$\begin{aligned}\Delta C_P &= 2(C_P)_{NH_3} - (C_P)_{N_2} - 3(C_P)_{H_2} \\ &= -9.92 - 2.3 \times 10^{-3} T + 10.2 \times 10^{-6} T^2; \\ \therefore \int_0^T \Delta C_P dT &= -9.92 T - 1.15 \times 10^{-3} T^2 + 3.4 \times 10^{-6} T^3.\end{aligned}$$

Insertion of this result in (60) gives

$$\Delta H = \Delta H_0 - 9.92 T - 1.15 \times 10^{-3} T^2 + 3.4 \times 10^{-6} T^3.$$

If ΔH is known at any one temperature T it is obviously possible to evaluate the constant ΔH_0 ; in this particular instance it is found to be $-19,000$ cal., so that

$$\Delta H = -19,000 - 9.92 T - 1.15 \times 10^{-3} T^2 + 3.4 \times 10^{-6} T^3,$$

and hence ΔH at any required temperature can be calculated. It should be noted that ΔH is given in calories, because the heat capacities are expressed in these units.

As is to be expected, the same result may be obtained by direct application of (64); using the heat capacities quoted above it follows that

$$\begin{aligned}\alpha &= 2(8.04) - \{6.5 + 3(6.5)\} = -9.92, \\ \beta &= 2(7 \times 10^{-4}) - \{10^{-3} + 3(9 \times 10^{-4})\} = -2.3 \times 10^{-3}, \\ \gamma &= 2(5.1 \times 10^{-6}) = 10.2 \times 10^{-6}.\end{aligned}$$

Insertion of these values in (64) then gives the same equation for ΔH as before.

THE SECOND LAW OF THERMODYNAMICS

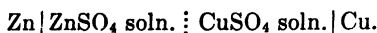
Spontaneous Processes—It was seen earlier that for a process to be thermodynamically reversible it must be carried out infinitesimally slowly, so that the system represents a continuous succession of equilibrium states. It is a matter of experience, however, that *all natural or spontaneous processes, i.e., processes occurring without external interference, are irreversible*; this is the basis of the **second law of thermodynamics**, and in fact the law has been stated in this form. A gas expands spontaneously into a vacuum, or into any region of lower pressure, until the pressure distribution is uniform, but a gas at uniform pressure is never observed to reverse the process. Similarly, heat is conducted spontaneously along a bar of metal hot at one end and cold at the other until the temperature is the same throughout; and heat will, in general, pass by radiation from a hotter to a colder body. The reverse of these changes is never found to occur spontaneously. Numerous other examples may be quoted of natural irreversible processes, e.g., diffusion of one gas into another, diffusion of a concentrated solution into water, production of heat by friction, and the conversion of chemical (internal) energy into heat in a chemical reaction.

Certain processes, such as expansion of a gas and dilution of a solution by diffusion, are irreversible when they occur spontaneously, but they can be carried out reversibly if required, at least in theory. For example, if the pres-

sure on the gas is reduced infinitesimally slowly, or if the solution is diluted in an analogous manner by the aid of a semipermeable membrane (cf. p. 652), the process is reversible. Some chemical reactions, such as that between metallic zinc and copper sulfate, viz.,



can also be made to take place reversibly; this can be achieved by setting up a galvanic cell, known as the Daniell cell, consisting of a zinc electrode in zinc sulfate solution separated by a porous partition from a copper electrode in copper sulfate solution, thus



If a very minute current is drawn from the cell the reaction shown above should occur in a reversible manner; on the other hand, if a minute current is passed through the cell, by applying an external E.M.F. which just exceeds that of the cell, the reaction is reversed (see p. 922). In practice the reversibility is not quite perfect, for various reasons, e.g., attack of the metal, interdiffusion of the solutions, etc., but the cell is at least theoretically reversible. If appreciable currents are drawn from or passed into the cell, heat is generated because of its internal electrical resistance, and the process is no longer reversible. Galvanic cells of the type under consideration are probably the nearest practical approach to completely reversible systems; the subject will be considered more fully in Chapter XII.

The Second Law of Thermodynamics.—Although the spontaneous processes described above never reverse themselves spontaneously, it is of interest to see how the original conditions could be re-established. By introducing a piston into the vessel, the gas which had expanded into a vacuum could be restored to its original volume by compression; work would have to be done on the gas, and at the same time an equivalent amount of heat would be produced and the temperature would rise. If the heat could be completely reconverted into work by means of a hypothetical machine, then the original state of the gas would have been restored and there would be no change in external bodies. It is a fundamental fact of experience, however, that complete conversion of heat into work is impossible, without resulting in some external change. For example, a cylinder of highly compressed gas placed in a heat reservoir will do work by pushing back a piston, and an exactly equivalent amount of heat will be absorbed from the reservoir. It will be seen, however, that the volume of the gas has increased, so that the complete conversion of heat into work is accompanied by a change in the system. If the gas is restored to its original state by compression, the work obtained in expansion will have to be done on the system and the equivalent amount of heat will be given up to the reservoir. The net result of the expansion and compression is that there is no heat absorbed and no work done. The failure in practice to convert heat completely into work leads to another way of expressing the second law of thermodynamics. It states that *it is impossible to construct a machine functioning in cycles which can con-*

vert heat completely into the equivalent amount of work without producing changes elsewhere. The term "functioning in cycles" is inserted to indicate that the machine must return exactly to its original state at regular stages, as explained below, so that it can operate continuously.

When a bar of metal, which was originally hot at one end, has attained a uniform temperature, the original condition could be restored in the following manner: heat is withdrawn from one end of the bar, converted completely into work, and then the work could be utilized to heat the other end of the metal bar by friction. Actually it is impossible to carry out this series of processes without leaving changes in external bodies for, as seen above, complete conversion of heat into work is impossible. Considerations of this kind help in understanding the form of the second law as enunciated by R. Clausius (1854): *heat cannot of itself, without the intervention of any external agency, pass from a colder to a hotter body.*

Perpetual Motion of the Second Kind.—Although perpetual motion of the first kind, i.e., the creation of energy, may be impracticable, there is a possibility that use could be made of the vast stores of energy in the earth and especially in the ocean; this has been called **perpetual motion of the second kind** (W. Ostwald). Imagine a ship at a somewhat lower temperature than the sea on which it floats; heat would pass to the ship where it could be utilized in a suitable machine to produce the work necessary for propulsion. As a result of the friction between the propeller and the water, the work could be reconverted into heat which would pass into the ocean, and the energy of the latter would remain unchanged. This suggestion is not contrary to the law of conservation of energy, but it is nevertheless found, as a result of experience, to be impossible. William Thomson, Lord Kelvin, (1852) has in fact stated "it is impossible to obtain work by cooling a body below the lowest temperature of the system," and this is another way of expressing the second law of thermodynamics. In order to convert heat continuously into work, it is necessary for the machine to take in heat at a higher temperature, referred to as the source, convert part of it into work, and give out the remainder at a lower temperature, called the sink. The conversion of heat into work by an engine working in cycles requires, therefore, an upper and a lower temperature; even then the conversion is only partial. It is now easy to understand why perpetual motion of the second kind is impossible. If the engine on a ship is at a lower temperature than the ocean it can take in heat, but in order to convert this into work part of the heat must be given up at a still lower temperature. There is, however, no means available for absorbing this heat, since all other bodies are at a higher temperature; it is evident therefore that the engine for converting the heat of the ocean into work cannot function.*

Cycles or Cyclic Processes.—A cycle is a process in which a system returns to its original state after a succession of changes; for example, the

* It should be made clear that the suggestion by G. Claude to utilize the energy of the tropical seas is not contrary to the second law of thermodynamics; he proposed to take advantage of the fact that the temperature of the surface water is considerably higher than that of the deeper levels. It should be possible, therefore, to take in heat from the water at the surface of the sea, convert part into work, and give out the remainder at a deeper level where the temperature is much lower.

two processes described on p. 185, and depicted in Fig. 34, constitute a cycle. The change in total energy is evidently zero, since after the completion of the cycle the system is restored to its initial condition, and so it follows from the first law that in a cyclic process the resultant work done by the system on its surroundings is equal to the net heat absorbed from the surroundings, that is $Q = W$; the algebraic sum of the heat terms is equal to the algebraic sum of the work terms.

Efficiency.—The impossibility in practice of converting heat completely into work shows that heat is different from other forms of energy. All other forms of energy are completely interconvertible, and all forms of energy may be completely changed into heat, but from the point of view of transformation into other kinds of energy heat is exceptional. The fraction of the heat absorbed by a machine which it can convert into work, or other energy, is called the **efficiency** of the machine. Thus if heat Q_2 is absorbed at the higher temperature T_2 and $-Q_1$ * is given out at the lower temperature T_1 , then according to the first law the difference $Q_2 + Q_1 = W$ is converted into work, and

$$\text{Efficiency of the machine} = \frac{Q_2 + Q_1}{Q_2} = \frac{W}{Q_2}. \quad (66)$$

It has been seen that when an isothermal change is carried out reversibly it leads to the maximum work of expansion; it follows, therefore, that a reversible machine will give the maximum efficiency, and the method of determining the efficiency will now be considered.

Carnot's Theorem.—An important simplification of the problem resulted from the theorem put forward by S. Carnot (1824) which may be stated in the following form. *All periodic machines working reversibly between the same temperatures of source and sink have the same efficiency.* That is to say, provided the machine functions reversibly its efficiency is independent of the nature of the substance, or substances, employed or the mode of operation. To prove this theorem, it may be imagined that two reversible machines, working between the same temperatures, have different efficiencies. Suppose that in each cycle machine I can convert a portion W of the heat Q_2 absorbed at T_2 , and give out the remainder, $Q_2 - W = Q_1$, at the lower temperature T_1 . On the other hand, in each cycle machine II is able to convert a smaller portion W' of the heat Q_2 absorbed at T_2 into work and give up a larger quantity, $Q_2 - W' = Q_1'$, at T_1 . Let the two machines be coupled together so that machine I operates in the direct manner whereas machine II functions in the reverse direction; this is permissible since the machines are assumed to be reversible. The various heat and work changes in each complete cycle are

* The convention stated on p. 187 is retained; heat absorbed by the system is positive, and heat given out is negative.

shown below.

I	II
Heat transferred at $T_2 = Q_2$	Heat transferred at $T_2 = -Q_2$
Work done $= W$	Work done $= -W'$
Heat transferred at $T_1 = -Q_1$	Heat transferred at $T_1 = Q'_1$

The net result for the complete cycle by the two machines, bringing them both back to their original state, is

$$\text{Heat absorbed at } T_1 = Q'_1 - Q_1 \quad \text{Work done} = W - W'.$$

Since Q_1 is equal to $Q_2 - W$, and Q' is equal to $Q_2 - W'$, it follows that the heat absorbed at T_1 is equal to the work done; that is to say, the combined machines functioning in cycles, can convert heat absorbed *at one temperature* completely into work. This is known to be impossible, and so it is a direct consequence of the second law of thermodynamics that the two reversible machines cannot have different efficiencies; the Carnot theorem follows, therefore, from the law.

Carnot's Cycle.—Since all reversible machines must have the same efficiency it is necessary to consider only one such machine; the one which lends itself to simple theoretical treatment is the cycle described by S. Carnot (1824). In this machine the working substance is 1 mole of an ideal gas, and it is contained in a cylinder fitted with a frictionless and weightless piston. It is supposed that there are available two very large heat reservoirs at T_2 and T_1 , and also that it is possible to surround the cylinder by a perfect nonconductor, thus permitting adiabatic processes to be carried out. The Carnot cycle involves four stages which can be represented diagrammatically (Fig. 37), the pressures and volumes being plotted as ordinates and abscissae, respectively.

I. The cylinder containing the 1 mole of gas, occupying volume V_1 , is placed in the constant temperature reservoir at T_2 , and expanded isothermally and reversibly, that is, infinitesimally slowly, until the volume has increased to V_2 ; the path is represented by AB in Fig. 37. Since the process is carried out reversibly with an ideal gas, the work done (see p. 197) is given by

$$W_I = RT_2 \ln V_2/V_1. \quad (67)$$

According to the first law the heat absorbed Q_2 is equal to W_I , the work done, since there is no change in the internal energy [cf. equation (40)].

II. The cylinder of gas is removed from the heat reservoir and surrounded by the nonconducting jacket, and expanded adiabatically to volume V_3 , that is, from B to C . The work is done at the expense of the kinetic energy of the gas molecules. There is thus a fall of temperature from T_2 to T_1 , and the work done (see p. 199) is

$$W_{II} = C_V(T_2 - T_1). \quad (68)$$

III. The cylinder is now placed in the heat reservoir at T_1 and the gas is compressed isothermally and reversibly from C to D ; the final volume

is V_4 , so that the work done is represented by

$$W_{III} = RT_1 \ln V_4/V_3. \quad (69)$$

The quantity W_{III} is equal to $-Q_1$, the heat evolved at T_1 , and represents work done *on* the gas, since stage III is a compression; in accordance with convention work done by the system is always considered, and actually W_{III} as written in (69) is negative, since V_3 is greater than V_4 , so that it implies that work is done on the gas.

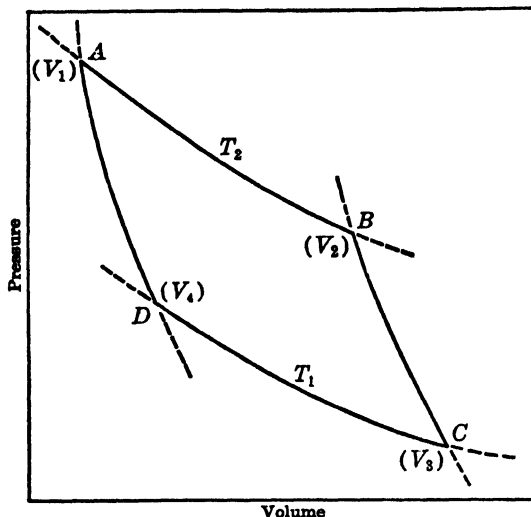


FIG. 37. Carnot's cycle

IV. The nonconducting jacket is replaced and the gas compressed adiabatically along DA , that is, until the initial state A is regained; the position of D is, of course, chosen so that it lies on the same adiabatic curve as A . The temperature of the gas rises from T_1 to T_2 , and hence

$$W_{IV} = C_V(T_1 - T_2), \quad (70)$$

that is, work is done *on* the gas.

As a result of these four stages a reversible cycle has been completed, and the total work done W is the sum of W_I , W_{II} , W_{III} and W_{IV} ; since W_{II} and W_{IV} , as given by (68) and (70), are equal but of opposite sign, it follows that

$$W = RT_2 \ln V_2/V_1 + RT_1 \ln V_4/V_3. \quad (71)$$

The net heat absorbed Q is equal to $Q_2 + Q_1$, and according to the first law of thermodynamics this must equal W , since in a complete cycle the total work done is equal to the net heat absorbed (p. 218); that is,

$$W = Q_2 + Q_1. \quad (72)$$

Since V_1 and V_4 lie on one adiabatic curve, and V_2 and V_3 are on another, it follows from (49) that

$$(V_4/V_1)^{\gamma-1} = T_2/T_1 \text{ and } (V_3/V_2)^{\gamma-1} = T_2/T_1;$$

$$\therefore V_4/V_1 = V_3/V_2 \text{ and } V_2/V_1 = V_3/V_4.$$

On substituting in (71) it is seen that

$$W = RT_2 \ln V_2/V_1 - RT_1 \ln V_2/V_1$$

$$= R(T_2 - T_1) \ln V_2/V_1. \quad (73)$$

Since Q_2 is equal to W_1 , as shown above, it follows, from (67), that

$$Q_2 = RT_2 \ln V_2/V_1, \quad (74)$$

and hence if (73) is divided by (74) there results

$$\frac{W}{Q_2} = \frac{T_2 - T_1}{T_2}. \quad (75)$$

Finally, if W is replaced by $Q_2 + Q_1$, as given by (72),

$$\frac{Q_2 + Q_1}{Q_2} = \frac{T_2 - T_1}{T_2}. \quad (76)$$

This relationship gives the efficiency of the reversible cycle under discussion, and so it represents the efficiency of any reversible machine operating between the temperatures T_2 and T_1 ; as is to be expected, the result is independent of the nature of the working substance. The efficiency is determined by the temperatures of the source and sink; only if the latter were at the absolute zero would it be possible to obtain an efficiency of unity, and so convert heat completely and continuously into work.

Thermodynamic Temperature Scale.—The establishment of a practical temperature scale generally depends on the thermal expansion of a particular substance, e.g., mercury or a gas; since the coefficients of expansion are not constant over a range of temperature, the scale defined in this manner will be subject to some uncertainty. In order to overcome this difficulty, W. Thomson (Lord Kelvin) suggested the **thermodynamic temperature scale** based on the efficiency of a reversible machine. Suppose there are two constant temperature heat reservoirs and a reversible machine operates between them, one acting as source and the other as sink; the temperature of each reservoir on the new scale is then defined as proportional to the quantity of heat transferred to or from it in the reversible cycle. If Q_2 is the heat transfer in the reservoir at the higher temperature and $-Q_1$ that at the lower temperature, then the respective temperatures on the thermodynamic, or Kelvin, scale are θ_2 and θ_1 , given by

$$-Q_2/Q_1 = \theta_2/\theta_1. \quad (77)$$

In this way the ratio of two temperatures is defined in a manner independent of the nature of any particular thermometric substance. The position of the

zero on the Kelvin scale may be derived in the following manner; invert each side of (77) and subtract the result from unity, thus

$$\frac{Q_2 + Q_1}{Q_2} = \frac{\theta_2 - \theta_1}{\theta_2}, \quad (78)$$

and putting $\theta_1 = 0$, that is, the zero of the scale, it follows that

$$\frac{Q_2 + Q_1}{Q_2} = 1. \quad (79)$$

The zero of the Kelvin scale is therefore the temperature of the sink for a reversible machine with an efficiency of unity, that is, one capable of converting heat completely into work. As already seen this is only possible at the absolute zero on the ideal gas scale of temperature; from this fact, and the identity of (76) and (78), it follows that the Kelvin scale and the gas scale, provided the gas is ideal, are really the same. In order that the two scales should be quite identical, it is necessary to define the size of the degree so as to be the same on both; for this definition reference may be made to p. 192. Because of the identity of the thermodynamic scale with that based on a hypothetical ideal gas, temperatures on the absolute scale, in which the melting point of ice is taken as 273.16° , are often referred to as "degrees Kelvin," and represented by the symbol $^\circ\text{K}$.

Entropy.—For a reversible Carnot cycle working between the absolute temperatures T_2 and T_1 , it has been seen that

$$\begin{aligned} \frac{Q_2 + Q_1}{Q_2} &= \frac{T_2 - T_1}{T_2}; \\ \therefore \frac{Q_2}{T_2} + \frac{Q_1}{T_1} &= 0, \end{aligned} \quad (80)$$

where Q_2 is the heat absorbed reversibly by the system at the temperature T_2 , and Q_1 is that absorbed, i.e., $-Q_1$ is the heat evolved, at the temperature T_1 .

Any reversible cycle may be regarded as being made up of a number of Carnot cycles; consider, for example, the cycle represented in Fig. 38 by the closed curve ABA . Imagine a series of isothermal and adiabatic curves drawn across the diagram, thus dividing it into a number of Carnot cycles. Starting from the point A and following all the cycles down successively and up back again to A , it can be shown that all the paths inside the figure cancel one another out, leaving only the zigzag outline. The larger the number of Carnot cycles taken in this manner the closer will the resultant path correspond to ABA , which is that of the reversible cycle under consideration; the latter may, therefore, be regarded as equivalent to an infinite number of Carnot cycles performed in succession. For each such cycle (80) shows that the sum of the Q/T terms involved is zero; hence, for all the Carnot cycles equivalent to the reversible cycle ABA , it follows that

$$\sum \frac{Q}{T} = 0, \quad (81)$$

where the symbol \sum is used to represent the summation of all the appropriate terms. For an infinite number of small cycles it is possible to write

$$\sum \frac{q}{T} = 0, \quad (82)$$

where q now represents an infinitesimally small reversible heat absorption; the summation is carried over all the isothermal stages in the complete

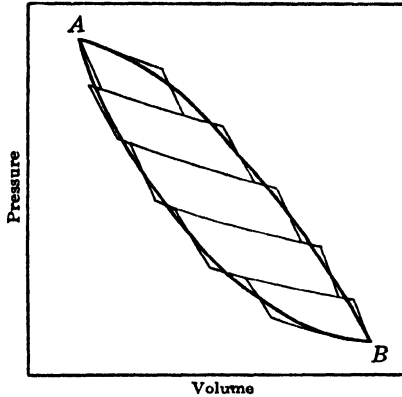


FIG. 38. Pressure-volume change as a succession of Carnot cycles

cycle. The sum in (82) may be divided into two parts, one for the path from A to B , and the other from B to A ; thus,

$$\sum \frac{q}{T} = \sum_{A \rightarrow B} \frac{q}{T} + \sum_{B \rightarrow A} \frac{q}{T} = 0. \quad (83)$$

It may be possible to go from A to B by a number of *different* reversible paths, always returning to A by the *same* reversible path BA . It is evident, therefore, from (83) that the value of $\sum_{B \rightarrow A} q/T$, and consequently of $\sum_{A \rightarrow B} q/T$, must be independent of the path between A and B , or between B and A , provided only it is reversible. The values of the sums are thus determined by the states A and B , that is, by the pressure, volume and temperature, and are independent of the path between them. It is thus possible to express the value of each sum in terms of a function S , which depends only on the initial and final states of the system viz.,

$$\sum_{A \rightarrow B} \frac{q}{T} = S_B - S_A = \Delta S, \quad (84)$$

S_A being the value of the function at A , and S_B that at B . The increase in the function accompanying the change from A to B is ΔS , and its value is given by the sum of the q/T terms between A and B . For an

infinitesimal stage of a reversible process, it follows, therefore, that

$$dS = \frac{q}{T}. \quad (85)$$

The quantity S is known as the **entropy** (Greek, *change*) of the system in a particular state. It is not easily described directly, and so is best defined in terms of the increase of entropy of a system; this is equal to the heat taken up *isothermally and reversibly* divided by the temperature at which it is absorbed. As seen above, the actual entropy of a system depends only on the state of that system, and consequently dS is an exact or complete differential.* As a consequence of this fact it is evident that if a system changes from A to B in an *irreversible* manner, then the change of entropy is still given by the sum of the q/T terms as in (84), where q is the heat that would be absorbed in each stage if the process were carried out *reversibly*. Since entropy is determined by a quantity of heat divided by the temperature, it is generally expressed in calories per degree; this is sometimes called the conventional "entropy unit," abbreviated to E.U.

Entropy of an Ideal Gas.—For an infinitesimal stage of a reversible change (19) is

$$q = \left(\frac{\partial E}{\partial V} \right)_T dV + C_v dT + w,$$

and for an ideal gas the term $(\partial E / \partial V)_T$ is zero, in accordance with definition. If the work done is against the external pressure, w is equal to PdV , and hence for an ideal gas

$$q = C_v dT + PdV. \quad (86)$$

Dividing by T , it follows that

$$dS = \frac{q}{T} = \frac{C_v}{T} dT + \frac{P}{T} dV.$$

For 1 mole of an ideal gas $PV = RT$, so that P may be replaced by RT/V , and C_v is now the molar heat capacity; hence,

$$dS = C_v \frac{dT}{T} + R \frac{dV}{V}. \quad (87)$$

On integration, C_v being assumed constant for an ideal gas, there is obtained

$$S = C_v \ln T + R \ln V + S_0, \quad (88)$$

where S_0 is the integration constant. For two states of the gas, represented by the subscripts 1 and 2, for the initial and final states, respec-

* It should be noted that although q is not a complete differential, q/T , equal to dS , is a complete differential, provided q is taken up reversibly.

tively, the entropies are

$$\begin{aligned} S_1 &= C_V \ln T_1 + R \ln V_1 + S_0, \\ S_2 &= C_V \ln T_2 + R \ln V_2 + S_0; \\ \therefore \Delta S &= S_2 - S_1 = C_V \ln T_2/T_1 + R \ln V_2/V_1. \end{aligned} \quad (89)$$

This equation gives the change in entropy for an appreciable change in the conditions of an ideal gas; at constant temperature the first term on the right hand side is zero, whereas at constant volume the second term disappears. Replacing V in (88) by RT/P , and utilizing the fact that $C_P - C_V = R$ for an ideal gas (p. 195), it can be readily deduced that

$$S = C_P \ln T - R \ln P + S'_0, \quad (90)$$

where S'_0 is equal to $S_0 + R \ln R$. It follows, therefore, that for a change from state 1 to state 2,

$$\Delta S = S_2 - S_1 = C_P \ln T_2/T_1 - R \ln P_2/P_1. \quad (91)$$

For an isothermal change, that is, the temperature is constant and hence T_1 and T_2 are equal, (89) and (91) give

$$\Delta S = R \ln V_2/V_1 = R \ln P_1/P_2. \quad (92)$$

In an expansion, $V_2 > V_1$ or $P_1 > P_2$, and so ΔS is positive, but for a contraction ΔS is negative; an isothermal expansion is, therefore, accompanied by an increase, and a contraction by a decrease, of entropy of the system.

In an adiabatic change q is zero, as no heat enters or leaves the system; such changes consequently do not result in any increase or decrease of entropy. They have therefore been called **isentropic changes**.

It can be readily shown, e.g., by the definition of dS as q/T , that entropy is an extensive property, in the sense described on page 186. Hence the entropy of n moles of an ideal gas is obtained upon multiplying by n any of the equations derived above for 1 mole. Further, the entropy of any system consisting of a mixture of gases, or other substances, is the sum of the individual entropies of the constituents, at the pressures or concentrations *existing in the mixture*. Thus, for a mixture of ideal gases, the entropy is given by the summation

$$S = \sum n(C_P \ln T - R \ln p + S'_0), \quad (93)$$

where n is the number of moles of each particular gas present in the mixture, and p is its partial pressure (see p. 302). It will be seen in Chapter IV that the partial pressure of an ideal gas is equal to xP , where x is the mole fraction (p. 303) of that particular gas in the mixture and P is the total pressure. Making this substitution, (93) becomes

$$S = \sum n(C_P \ln T - R \ln x - R \ln P + S'_0).$$

The difference between the entropy of the mixture and the sum of the entropies of the separate gases each at a pressure P , i.e.,

$$\sum n(C_P \ln T - R \ln P + S'_0),$$

is called the **entropy of mixing**; thus,

$$\Delta S_{\text{mix.}} = - \sum n_i R \ln x_i,$$

where n_i and x_i represent the number of moles and the mole fractions, respectively, of each constituent of the mixture. Since the sum of all the n_i 's is equal to the total number of moles, it can be readily shown that for a *total* of 1 mole, the entropy of mixing is given by

$$\Delta S_{\text{mix.}} = - \sum x_i R \ln x_i. \quad (94)$$

Since the x_i 's are all fractions, the entropy of mixing is always positive.

✓ **Entropy Changes at Constant Volume or Constant Pressure.**—When the pressure inside a system is virtually the same as the external pressure, as is usually the case in processes of physicochemical interest, a small expansion or compression will take place reversibly. If the work done is then only mechanical work against the external pressure, w in (6) can be replaced by PdV , so that

$$q = dE + PdV, \quad (95)$$

where q is the heat absorbed (reversibly) by the system from its surroundings. If this is divided by T , the value of the entropy change of the system is given by

$$dS = \frac{q}{T} = \frac{dE + PdV}{T}. \quad (96)$$

At constant volume this becomes $dS = dE/T$, or

$$\left(\frac{\partial S}{\partial E} \right)_v = \frac{1}{T}. \quad (97)$$

For any system, not necessarily an ideal gas, it has been shown (p. 189) that $C_V = (\partial E / \partial T)_V$, and so from (97) and (14) there follows

$$C_V = T \left(\frac{\partial S}{\partial T} \right)_V. \quad (98)$$

Integrating (98) between temperatures T_1 and T_2 , it is readily deduced that the change of entropy *at constant volume* resulting from a change of temperature from T_1 to T_2 is given by

$$\Delta S = \int_{T_1}^{T_2} C_V \frac{dT}{T} = \int_{T_1}^{T_2} C_V d \ln T. \quad (99)$$

Since $H = E + PV$, by definition (p. 189), it follows that at constant pressure

$$dH = dE + PdV$$

and so, from (96), under these conditions $dS = dH/T$, or

$$\left(\frac{\partial S}{\partial H} \right)_P = \frac{1}{T}. \quad (100)$$

It is known (p. 190) that for any system $C_P = (\partial H/\partial T)_P$, and so from (100),

$$C_P = T \left(\frac{\partial S}{\partial T} \right)_P. \quad (101)$$

For a change of temperature from T_1 to T_2 at constant pressure, the corresponding entropy change is given by

$$\Delta S = \int_{T_1}^{T_2} C_P \frac{dT}{T} = \int_{T_1}^{T_2} C_P d \ln T. \quad (102)$$

If C_V and C_P are taken as constant, e.g., for an ideal gas, then (99) and (102) can be integrated; the results are found to be identical with those obtained for changes at constant volume and pressure, respectively, from (89) and (91).

✓ **Entropy Changes in Irreversible Processes.**—It has been seen (p. 223) that in a reversible cycle

$$\sum_{A \rightarrow B} \frac{q}{T} + \sum_{B \rightarrow A} \frac{q}{T} = 0;$$

hence

$$\Delta S_{A \rightarrow B} + \Delta S_{B \rightarrow A} = 0.$$

That is to say, there is no entropy change of the system when a reversible cycle has been completed. Further, in any reversible process the heat quantity q taken up by the system is supplied in a reversible manner by a heat reservoir; hence the change in entropy of the surroundings will be equal to the change in entropy of the system, but opposite in sign since the latter takes in heat whereas the former gives it out. In every reversible process, therefore, the sum of the entropy changes of the system and its surroundings is zero, and hence in a complete cycle there is no resultant change of entropy.

The efficiency of an irreversible cycle must be less than that of a reversible one, since the latter operates under maximum work conditions (p. 218); it follows then that

$$\left(\frac{Q_2^* + Q_1^*}{Q_2^*} \right)_{irr.} < \frac{T_2 - T_1}{T_2} \quad (103)$$

and, therefore,

$$\frac{Q_2^*}{T_2} + \frac{Q_1^*}{T_1} < 0, \quad (104)$$

the asterisk being used to indicate irreversibility. It is not necessary for both isothermal stages of the cycle to be irreversible for (104) to hold; it is sufficient for either one of them to satisfy this condition. In general, for any cycle, involving a number of isothermal stages, the sum of the Q/T , or q/T , terms, will be less than zero if one or more of these stages is irreversible.

Consider a cycle ABA made up of a path $A \rightarrow B$, involving an ir-

reversible stage (or stages), and a reversible path $B \rightarrow A$; it follows from the foregoing arguments that

$$\sum_{A \rightarrow B} \frac{q^*}{T} + \sum_{B \rightarrow A} \frac{q}{T} < 0. \quad (105)$$

By definition [equation (84)],

$$S_A - S_B = \sum_{B \rightarrow A} \frac{q}{T},$$

where q is taken up reversibly; hence,

$$\sum_{A \rightarrow B} \frac{q^*}{T} + S_A - S_B < 0 \quad (106)$$

or

$$\sum_{A \rightarrow B} \frac{q^*}{T} < S_B - S_A. \quad (107)$$

In an irreversible process $A \rightarrow B$, therefore, the summation in (107) is actually less than the gain of entropy $S_B - S_A$ of the system. Now consider the surroundings after the process $A \rightarrow B$ has occurred; it has supplied a quantity of heat represented by $\sum_{A \rightarrow B} q^*$, and although this has been taken up by the system in an irreversible manner, the change in entropy of the surroundings is evaluated by supposing the process to be carried out reversibly (cf. p. 224). The decrease of entropy of the surroundings is consequently given by $\sum_{A \rightarrow B} q^*/T$. The total entropy change of the system and its surroundings in the process $A \rightarrow B$ is therefore obtained as follows:

$$\begin{aligned} \text{Gain of entropy of system} &= S_B - S_A, \\ \text{Decrease of entropy of surroundings} &= \sum_{A \rightarrow B} \frac{q^*}{T}, \\ \text{Net gain of entropy} &= (S_B - S_A) - \sum_{A \rightarrow B} \frac{q^*}{T}. \end{aligned} \quad (109)$$

Since the summation term is less than the $S_B - S_A$, the net gain of entropy is positive, and so it must be concluded that as the result of an irreversible process there is a positive net gain of entropy. If the cycle is completed by the reversible path $B \rightarrow A$, there is no resultant entropy change, and hence the complete cycle must lead to an increase of entropy of the system and the surroundings combined. It follows, therefore, that an irreversible process, or a cycle in which any part is irreversible, is accompanied by a gain of entropy when the system and its surroundings are considered. Since natural, or spontaneous, changes are irreversible, all such processes are accompanied by a net increase of entropy. From some points of view this is the most important consequence of the second law of thermodynamics. The two laws may thus be expressed, following

R. Clausius, in the somewhat analogous forms:

First Law: *The total amount of energy in Nature is constant*

Second Law: *The total amount of entropy in Nature is increasing.*

✓ **Entropy and Probability.**—The second law of thermodynamics is a statistical law which can only be applied to a system consisting of a large number of particles, as, for example, an assemblage of molecules. If it were possible to work with systems of single, or a few, molecules, the law might well fail. The probability that all the molecules in a gas will be equally spaced and move in an ordered manner relative to each other is very small, for as a result of impacts the distances of separation and velocities will vary continuously (cf. p. 248). In other words, the establishment of a condition of chaotic movement and of random distribution, which occurs quite spontaneously, is more probable than an ordered state. All spontaneous processes are accompanied by an increase of entropy; it is possible, therefore, to regard entropy as a measure of the "randomness" or "state of chaos" in a given system, and to suppose that all natural processes lead to an increase in random distribution. Since a disordered state is more probable than one of complete order, entropy and probability are evidently related, and for this purpose L. Boltzmann (1896) defined the **thermodynamic probability** of a system as the ratio of the probability of an actual state to one, having the same total energy and volume, in which the molecules are completely ordered. If S is the entropy and W the probability of a system, then it is possible to write $S = f(W)$. To ascertain the nature of the function, consider two systems having entropies S_1 and S_2 , and probabilities W_1 and W_2 . The entropy S of the combined system is $S_1 + S_2$, whereas its probability W is the product $W_1 \times W_2$, so that

$$S = S_1 + S_2 = f(W_1 W_2);$$

but since $S_1 = f(W_1)$ and $S_2 = f(W_2)$, it follows that

$$f(W_1 W_2) = f(W_1) + f(W_2).$$

To satisfy this condition it is obvious that the function must be logarithmic, and hence it is possible to write

$$S = k \ln W + \text{constant}. \quad (110)$$

The constant k was shown to be equal to the gas constant per single molecule, the Boltzmann constant, i.e., R/N , where N is the Avogadro number. If it is assumed that the entropy depends only on the thermodynamic probability of the system (M. Planck, 1912), defined in the proper manner, (110) becomes

$$S = k \ln W. \quad (111)$$

This important equation has become the basis of modern statistical thermodynamics, to which reference will be made in Chapter XI.

THERMODYNAMIC FUNCTIONS

Free Energy and Work Functions.—The quantities E , H and S are dependent only on the state of the system, and it is convenient to define

two further functions with the same characteristic property, viz.,

$$A = E - TS \quad (112)$$

$$F = H - TS. \quad (113)$$

Like E , H and S , the functions A and F are also extensive properties. For an *isothermal process*, these definitions may be written in the form

$$\Delta A = \Delta E - T\Delta S \quad (112a)$$

$$\Delta F = \Delta H - T\Delta S, \quad (113a)$$

and upon subtraction it is seen that

$$\Delta F - \Delta A = \Delta H - \Delta E. \quad (114)$$

If ΔH is measured at constant pressure, then by (11), $\Delta H - \Delta E = P\Delta V$, and hence (114) gives

$$\Delta F = \Delta A + P\Delta V. \quad (115)$$

The quantity A was called the "free energy" by H. von Helmholtz, but it is now usually referred to as the **work function**, for the following reason. If ΔS in (112a) is replaced by Q/T , where Q is the heat taken up reversibly, it is seen that for an isothermal process $\Delta A = \Delta E - Q$; comparison with (5) then shows that $-\Delta A$ is equal to W . The decrease in A is thus a measure of the maximum (reversible) work which can be made available in the given isothermal process. It should be noted that W will include *all* forms of work, viz., mechanical work against the external pressure, electrical work, etc.

The function F is variously called the thermodynamic potential, the Gibbs function, or, more commonly at the present time, the **free energy** (G. N. Lewis and M. Randall). Since ΔA is equal to $-W$, as just shown, it follows from (115) that $-\Delta F$ is equal to $W - P\Delta V$, where $P\Delta V$ is the work done in expansion at constant pressure in the course of the process. The decrease of free energy in a given isothermal process at constant pressure is thus equal to the *net work*, i.e., the work other than that due to a volume change, which the system is capable of performing.

Differentiation of (112) gives

$$dA = dE - TdS - SdT, \quad (116)$$

and by (96), for a reversible process in which the work done is only that due to a change of volume, (116) becomes, by substitution for TdS ,

$$dA = -PdV - SdT. \quad (117)$$

At constant volume, dV is zero, and so (117) reduces to

$$\left(\frac{\partial A}{\partial T}\right)_v = -S. \quad (118)$$

Since $H = E + PV$, by definition, differentiation of (113) gives

$$dF = dE + PdV + VdP - TdS - SdT. \quad (119)$$

Again, assuming the process to be reversible and supposing that the work is all due to expansion, this becomes

$$dF = VdP - SdT. \quad (120)$$

At constant pressure, dP is zero, and hence

$$\left(\frac{\partial F}{\partial T}\right)_P = -S. \quad (121)$$

It should be emphasized that in deriving the foregoing equations it has been assumed, apart from the condition concerning the work of expansion, that the amount and chemical composition of the system remains unchanged. Such changes would, of course, affect the values of the extensive properties, viz., E , H , A , F , etc. In other words, the results are applicable to what is called a **closed system**. For example, in a system consisting of several phases (cf. p. 477), matter may pass from one phase to another; hence each phase is an open system, although the whole system may be closed, since no matter leaves or enters it from outside. The equations applicable to open systems will be considered later (p. 237).

The Gibbs-Helmholtz Equation.—Suppose A_1 is the work function for a system in the initial state and at temperature T , and let the corresponding value at $T + dT$ be $A_1 + dA_1$; similarly for the final state of the system let the work functions be A_2 and $A_2 + dA_2$ for the two temperatures. From (118) it is seen that dA and dT are connected at constant volume by the relationship

$$dA = -SdT \quad (118a)$$

and so,

$$dA_1 = -S_1dT \quad \text{and} \quad dA_2 = -S_2dT,$$

where S_1 and S_2 are the entropies of the system in the initial and final states. Subtraction of dA_1 from dA_2 gives

$$d(A_2 - A_1) = -(S_2 - S_1)dT,$$

and writing ΔA for $A_2 - A_1$, the increase in the work function, and ΔS for $S_2 - S_1$, the increase of entropy, it follows that

$$d(\Delta A) = -\Delta SdT, \quad (122)$$

or, remembering that constant volume has been assumed,

$$\left(\frac{\partial(\Delta A)}{\partial T}\right)_V = -\Delta S. \quad (123)$$

Introducing this value for ΔS into (112a), there results

$$\Delta A - \Delta E = T \left(\frac{\partial(\Delta A)}{\partial T} \right)_v \quad (124)$$

This is one form of the equation deduced by J. Willard Gibbs (1875) and by H. von Helmholtz (1882).

Since under constant pressure conditions (121) gives

$$dF = -SdT, \quad (121a)$$

it can be deduced in a manner exactly analogous to that employed above that,

$$d(F_2 - F_1) = -(S_2 - S_1)dT,$$

from which it follows that

$$\left(\frac{\partial(\Delta F)}{\partial T} \right)_p = -\Delta S, \quad (123a)$$

and hence introduction into (113a) gives the alternative form of the Gibbs-Helmholtz equation

$$\Delta F - \Delta H = T \left(\frac{\partial(\Delta F)}{\partial T} \right)_p. \quad (125)$$

Although the forms of the Gibbs-Helmholtz equation have been derived from expressions applicable to reversible changes, the final results, viz., (124) and (125), hold for any process, reversible or irreversible, since the values of ΔA , ΔE , ΔF and ΔH depend only on the initial and final states of the system.

The importance of the Gibbs-Helmholtz equation lies in the fact that it permits the calculation of ΔE or ΔH , provided ΔA or ΔF , and their respective temperature coefficients, are known. By means of the Gibbs-Helmholtz equation the heat change in a reaction may be calculated from the E.M.F.'s of galvanic cells (p. 924).

Work Function and Free Energy in Isothermal Changes.—Under isothermal conditions dT is zero, and so for a reversible process in which the work is due to expansion only, (117) becomes

$$dA = -PdV \quad (126)$$

and similarly, from (120)

$$dF = VdP. \quad (127)$$

For an appreciable isothermal change between the initial and final states represented by the subscripts 1 and 2, respectively, it follows, therefore, that

$$\Delta A = \int_{V_1}^{V_2} dA = - \int_{V_1}^{V_2} PdV \quad (128)$$

and

$$\Delta F = \int_{F_1}^{F_2} dF = \int_{P_1}^{P_2} V dP. \quad (129)$$

These equations hold for reversible and irreversible process, provided the P and V values apply to equilibrium, i.e., reversible, conditions for the particular (constant) temperature.

To be able to integrate (128) and (129) it is necessary to have an equation connecting P and V ; this is possible, for example, for an ideal gas, where for 1 mole $PV = RT$. It follows then that

$$\Delta A = RT \ln \frac{V_2}{V_1} \quad \text{and} \quad \Delta F = RT \ln \frac{P_2}{P_1}. \quad (130)$$

Since V_1/V_2 is equal to P_2/P_1 for an ideal gas it follows that ΔA and ΔF are equal. It is important to point out that this equality only applies to an ideal gas, but is not true generally; considerable confusion has arisen because of the failure to realize this limitation.

Conditions of Equilibrium in Closed Systems.—It was seen on p. 228 that in an irreversible, e.g., spontaneous, process the sum of the q/T terms was actually less than the gain in entropy; for an infinitesimal stage of a spontaneous change, therefore, $dS > q/T$. Since the heat absorbed by the system, i.e., q , is equal to $dE + PdV$, assuming the work w to be only that due to a volume change, it follows that for a spontaneous process

$$dS > \frac{dE + PdV}{T}.$$

On the other hand, for a reversible process, (96) gives

$$dS = \frac{dE + PdV}{T},$$

so that it is possible to combine the two results in the form

$$dS \geq \frac{dE + PdV}{T}, \quad (131)$$

where the "greater than" sign refers to a spontaneous (irreversible) process and the "equals" sign to a reversible one. If E and V are taken as constant, so that dE and dV are zero, it follows that

$$(\partial S)_{E,V} \geq 0. \quad (132)$$

Since a reversible process involves a succession of equilibrium states, (132) may be used as a criterion of equilibrium or of a spontaneous process. It is evident from (132), therefore, that the entropy of a system will be a maximum at equilibrium, while a spontaneous process, which represents an approach toward equilibrium, is accompanied by an increase of entropy, at constant energy and volume.

By substitution for TdS in (116) the alternatives derived from (131), viz., $TdS \geq dE + PdV$, it is seen that

$$dA \leq -PdV - SdT,$$

so that at constant temperature and volume,

$$(\partial A)_{T,V} \leq 0, \quad (133)$$

where the "less than" sign now refers to a spontaneous process while the "equals" applies to a state of equilibrium.

Similar substitution in (119) gives

$$dF \leq VdP - SdT,$$

and hence at constant temperature and pressure,

$$(\partial F)_{T,P} \leq 0. \quad (134)$$

For an appreciable process taking place under these conditions, it is possible to express (134) as

$$(\Delta F)_{T,P} \leq 0. \quad (134a)$$

It follows from these results that a spontaneous process, at constant temperature and pressure, must be accompanied by a decrease in the free energy of the (closed) system, while at equilibrium the value is a minimum. The fact that the free energy of a system is a minimum at equilibrium, is of fundamental importance in physical chemistry. Spontaneous processes will always occur in the direction of a decrease of free energy, and the diminution of the free energy in any particular process may be taken as a measure of the tendency for that process to occur. Thomsen and Berthelot (p. 211) suggested that the heat evolved in a reaction was a measure of this tendency but their view was shown to be unsatisfactory, especially for chemically reversible reactions. This objection does not apply to the free energy, since it is a minimum at equilibrium, and there will be a decrease no matter from which direction the equilibrium is approached. The subjects of free energy and equilibrium will be treated more fully from these points of view in Chapter XI.

✓ **Heat Capacities at Constant Pressure and Constant Volume.**—The use of thermodynamic functions permits the development of an equation connecting the heat capacities at constant pressure and constant volume for any substance, not necessarily a gas. Since the entropy of a given amount of a substance depends only on the state of the system, that is on two of the three variables, pressure, temperature and volume, dS is a complete differential, as stated above, and so it is possible to write

$$dS = \left(\frac{\partial S}{\partial T} \right)_P dT + \left(\frac{\partial S}{\partial P} \right)_T dP.$$

Upon dividing by dT , and stipulating constant volume (see footnote, p. 194),

the result is

$$\left(\frac{\partial S}{\partial T}\right)_V = \left(\frac{\partial S}{\partial T}\right)_P + \left(\frac{\partial S}{\partial P}\right)_T \left(\frac{\partial P}{\partial T}\right)_V. \quad (135)$$

It has been seen [equations (98) and (101)], that

$$C_V = T \left(\frac{\partial S}{\partial T}\right)_V \quad \text{and} \quad C_P = T \left(\frac{\partial S}{\partial T}\right)_P,$$

and if these are inserted in (135) there is obtained

$$C_P - C_V = -T \left(\frac{\partial S}{\partial P}\right)_T \left(\frac{\partial P}{\partial T}\right)_V. \quad (136)$$

Further, by (120), $dF = VdP - SdT$, and by the rules of partial differentiation it follows that

$$\left(\frac{\partial V}{\partial T}\right)_P = - \left(\frac{\partial S}{\partial P}\right)_T. \quad (137)$$

Since pressure, volume and temperature are related,

$$\left(\frac{\partial V}{\partial P}\right)_T \left(\frac{\partial P}{\partial T}\right)_V \left(\frac{\partial T}{\partial V}\right)_P = -1, \quad (138)$$

and so from (136), (137) and (138),

$$C_P - C_V = -T \left(\frac{\partial V}{\partial T}\right)_P^2 / \left(\frac{\partial V}{\partial P}\right)_T. \quad (139)$$

The quantity $\frac{1}{V} \left(\frac{\partial V}{\partial T}\right)_P$ is equal to the coefficient of thermal expansion α_v of the substance, V being the volume at the temperature T , and $-\frac{1}{V} \left(\frac{\partial V}{\partial P}\right)_T$ is the compressibility coefficient β ; hence,

$$C_P - C_V = TV\alpha_v^2/\beta. \quad (140)$$

This equation applies equally to solids, liquids and gases, the volume V referring to the same quantity of substance as do the heat capacities, e.g., 1 g. atom or 1 mole. It is readily shown for an ideal gas, since $PV = RT$ for 1 mole, that α_v is equal to R/PV while β is RT/P^2V , and hence (140) reduces to the familiar form

$$C_P - C_V = R. \quad (141)$$

✓ **Thermodynamic Equations of State.**—The equations derived above may also be utilized to develop equations of state, giving the relationship between pressure, volume and temperature for all states of matter; they are particularly useful, however, in the study of the gaseous and liquid states (pp. 287, 479). As seen above [equation (95)], for a reversible change in which the work is due to expansion only $q = dE + PdV$, and since q is equal to TdS ,

$$dE = TdS - PdV, \quad (142)$$

and hence it follows that

$$\left(\frac{\partial E}{\partial V}\right)_T = T \left(\frac{\partial S}{\partial V}\right)_T - P. \quad (143)$$

By the rules of partial differentiation, it follows from (117) that

$$\left(\frac{\partial S}{\partial V}\right)_T = \left(\frac{\partial P}{\partial T}\right)_V, \quad (144)$$

and substitution in (143) leads to

$$\left(\frac{\partial E}{\partial V}\right)_T = T \left(\frac{\partial P}{\partial T}\right)_V - P, \quad (145)$$

which is a thermodynamic equation of state. For an ideal gas $(\partial E/\partial V)_T$ is zero, as defined previously, and so it follows that at constant volume, dP/P is equal to dT/T ; that is to say, the pressure is proportional to the absolute temperature. If now Boyle's law is assumed, it follows that at constant pressure the volume must be proportional to the absolute temperature, as required by Gay-Lussac's law. The latter is thus a thermodynamic necessity for an ideal gas to which Boyle's law is applicable.

Another thermodynamic equation of state may be deduced as follows. Differentiate the equation $H = E + PV$ which defines H (p. 189); thus,

$$dH = dE + PdV + VdP, \quad (146)$$

and substitute in (142), so that

$$dH = TdS + VdP. \quad (147)$$

Dividing by dP , at constant temperature, then

$$\left(\frac{\partial H}{\partial P}\right)_T = T \left(\frac{\partial S}{\partial P}\right)_T + V. \quad (148)$$

It is now possible to introduce (137) obtained by partial differentiation from (120); thus,

$$\left(\frac{\partial V}{\partial T}\right)_P = - \left(\frac{\partial S}{\partial P}\right)_T,$$

and substitution in (148) gives

$$V = T \left(\frac{\partial V}{\partial T}\right)_P + \left(\frac{\partial H}{\partial P}\right)_T. \quad (149)$$

Another thermodynamic equation for the difference between the heat capacities of any substance at constant pressure and volume may be obtained by combination of (32) and (145); there is obtained immediately

$$C_P - C_V = T \left(\frac{\partial P}{\partial T}\right)_V \left(\frac{\partial V}{\partial T}\right)_P. \quad (150)$$

This equation must be regarded as merely an alternative form of (32) and (139).

Partial Molar Quantities.—The magnitude of an extensive property, such as E , F , S , etc., is dependent on the amounts of the constituents of the system, as well as upon its state, e.g., temperature and pressure. Hence, for an **open system**, i.e., one to and from which the transfer of matter can occur, possible changes in the amount of material must be taken into consideration. If X represents any extensive property of the type described above then for a small change in the system, dX , being a complete differential, should be expressed as

$$dX = \left(\frac{\partial X}{\partial T} \right)_{P, n_1, n_2, \dots} dT + \left(\frac{\partial X}{\partial P} \right)_{T, n_1, n_2, \dots} dP + \left(\frac{\partial X}{\partial n_1} \right)_{T, P, n_2, \dots} dn_1 + \left(\frac{\partial X}{\partial n_2} \right)_{T, P, n_1, \dots} dn_2 + \dots, \quad (151)$$

where n_1, n_2, n_3, \dots , are the numbers of moles of the various constituents, represented by 1, 2, 3, \dots , making up the whole system. The quantities $\partial X / \partial n_i$, called the corresponding **partial molar quantities**, are represented by writing a bar over the symbol for the particular property; thus,

$$\left(\frac{\partial X}{\partial n_1} \right)_{T, P, n_2, \dots} = \bar{X}_1; \quad \left(\frac{\partial X}{\partial n_2} \right)_{T, P, n_1, \dots} = \bar{X}_2, \text{ etc.} \quad (152)$$

For example, if X represents the volume of the system, then $\partial V / \partial n_1$, written \bar{V}_1 , would be called the partial molar volume of the constituent 1; similarly, it is possible to have partial molar free energy, partial molar heat content, etc. The physical significance of the partial molar quantity for a given substance is seen to be the increase in a property of the system, e.g., volume, resulting from the addition at constant temperature and pressure of 1 mole of that particular substance to such a large quantity of the system that there is no appreciable change in the concentration. Using the abbreviation proposed, (151) may be written

$$dX = \left(\frac{\partial X}{\partial T} \right)_{P, n_1, n_2, \dots} dT + \left(\frac{\partial X}{\partial P} \right)_{T, n_1, n_2, \dots} dP + \bar{X}_1 dn_1 + \bar{X}_2 dn_2 + \dots \quad (153)$$

For a system at constant temperature and pressure this simplifies to

$$dX = \bar{X}_1 dn_1 + \bar{X}_2 dn_2 + \dots, \quad (154)$$

which on integration gives

$$X = n_1 \bar{X}_1 + n_2 \bar{X}_2 + \dots \quad (155)$$

From this equation it is seen that \bar{X} may be regarded as the contribution per mole of each particular constituent to the total value of the property X in the system under consideration. The magnitude of \bar{X} may, of course, vary with the concentration of the particular constituent in the system.

By the general differentiation of (155),

$$dX = (n_1 d\bar{X}_1 + \bar{X}_1 dn_1) + (n_2 d\bar{X}_2 + \bar{X}_2 dn_2) + \dots \quad (156)$$

$$= (n_1 d\bar{X}_1 + n_2 d\bar{X}_2 + \dots) + (\bar{X}_1 dn_1 + \bar{X}_2 dn_2 + \dots). \quad (157)$$

Comparing this result with (154) it follows that at constant temperature and pressure

$$n_1 d\bar{X}_1 + n_2 d\bar{X}_2 + \cdots = 0. \quad (158)$$

In a mixture containing only two constituents,

$$n_1 d\bar{X}_1 = -n_2 d\bar{X}_2, \quad (159)$$

and this simple relationship is the basis of the important Gibbs-Duhem equation, first deduced by J. Willard Gibbs (1875) and later, independently, by P. Duhem (1886).

The Chemical Potential.—The partial molar free energy as defined above is a property frequently employed in thermodynamics; it can be represented by the symbol \bar{F} , following the method of Lewis and Randall. It is, however, identical with the quantity known as the **chemical potential** described by Gibbs, and to which he gave the symbol μ ; as this method of representation and the term chemical potential are being increasingly employed in the literature of physical chemistry, they will be adopted here. According to the definitions given above

$$\left(\frac{\partial F}{\partial n_1} \right)_{T, P, n_2, n_3, \dots} = \bar{F}_1 = \mu_1 \quad (160)$$

and, therefore, rewriting (151) with the appropriate symbols

$$dF = \left(\frac{\partial F}{\partial T} \right)_{P, n_1, n_2, \dots} dT + \left(\frac{\partial F}{\partial P} \right)_{T, n_1, n_2, \dots} dP + \mu_1 dn_1 + \mu_2 dn_2 + \cdots \quad (161)$$

If there is no alteration in the number of molecules of the various constituents as a result of a change in the system, that is to say, the system is a closed one, then dn_1, dn_2, \dots , are all zero, and hence

$$dF = \left(\frac{\partial F}{\partial T} \right)_{P, n_1, n_2, \dots} dT + \left(\frac{\partial F}{\partial P} \right)_{T, n_1, n_2, \dots} dP. \quad (162)$$

It has been previously shown [equation (120)] that for such a change,

$$dF = VdP - SdT, \quad (120)$$

and, therefore, equating coefficients in (120) and (162) it is seen that

$$\left(\frac{\partial F}{\partial T} \right)_{P, n_1, n_2, \dots} = -S, \quad (163)$$

which is identical with (121), as it should be, and

$$\left(\frac{\partial F}{\partial P} \right)_{T, n_1, n_2, \dots} = V. \quad (164)$$

The same result for a closed system can be obtained directly from (120).

Condition of Equilibrium.—It is evident that (161) can now be written

$$dF = -SdT + VdP + \mu_1 dn_1 + \mu_2 dn_2 + \cdots, \quad (165)$$

and for a process at constant temperature and pressure

$$(\partial F)_{T, P} = \mu_1 dn_1 + \mu_2 dn_2 + \dots \quad (166)$$

One of the conditions for a closed system at equilibrium (p. 234) is that $(\partial F)_{T, P}$ is zero; hence, for equilibrium, at constant temperature and pressure, in such a system

$$\mu_1 dn_1 + \mu_2 dn_2 + \dots = 0 \quad \text{or} \quad \Sigma \mu dn = 0. \quad (167)$$

Partial Molar Volume.—By simple calculus it is possible to relate the partial molar volume of the constituent of a homogeneous mixture to other properties in a manner which will be found later to be useful. If (160) is differentiated with respect to P , and (164) with respect to n_1 , the results should be equal; thus,

$$\left(\frac{\partial \mu_1}{\partial P} \right)_{T, n_1, n_2, \dots} = \left(\frac{\partial V}{\partial n_1} \right)_{P, T, n_2, \dots} \quad (168)$$

The right hand side of this equation is, by definition, equal to the partial molar volume \bar{V}_1 , so that

$$\bar{V}_1 = \left(\frac{\partial \mu_1}{\partial P} \right)_{T, n_1, n_2, \dots} \quad (169)$$

The change of chemical potential of a given constituent with pressure, at constant temperature and composition, is therefore equal to the partial molar volume of that constituent.

Determination of Partial Molar Quantities.—Graphical methods are generally employed for evaluating partial molar quantities; some of these will be described briefly for systems of two components. An obvious procedure is to plot the value of the extensive property X , at constant temperature and pressure, for various mixtures of the two components, against the number of moles n_1 of one of them, the value of n_2 being kept constant; the slope of the curve at any point gives \bar{X}_1 at that concentration. Owing to the difficulty of drawing the tangents exactly this method is not accurate. A second method involves the use of the **apparent molar property** ϕ . If X is the value of a particular property for a mixture of n_1 moles of one component and n_2 moles of the other, and X_1^0 is the value of the property *per mole of pure component* (1), then the apparent molar property ϕ_2 for the component (2) is given by $(X - n_1 X_1^0)/n_2$; hence,

$$n_2 \phi_2 = X - n_1 X_1^0. \quad (170)$$

Keep n_1 constant, so that $n_1 X_1^0$ is constant, and differentiate with respect to n_2 ; thus,

$$n_2 \left(\frac{\partial \phi_2}{\partial n_2} \right)_{n_1} + \phi_2 = \left(\frac{\partial X}{\partial n_2} \right)_{n_1} = \bar{X}_2 \quad (171)$$

or

$$\left(\frac{\partial \phi_2}{\partial \ln n_2} \right)_{n_1} + \phi_2 = \bar{X}_2. \quad (171a)$$

If ϕ_2 is determined for various values of n_2 , from a knowledge of X and X_1^0 , then the partial molar quantity \bar{X}_2 may be obtained by plotting ϕ_2 against $\ln n_2$ for constant values of n_1 . Similarly \bar{X}_1 can be determined from the plot of ϕ_1 against $\ln n_1$.

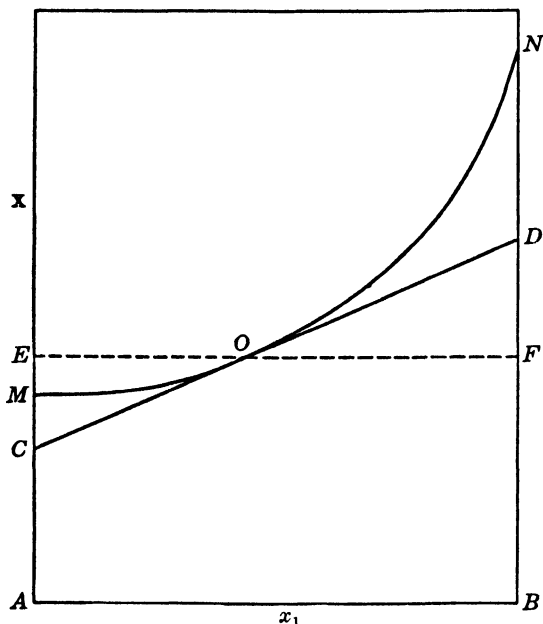


FIG. 39. Plot of the average value per mole of a property (x) against the mole fraction (x_1)

For many purposes the method of intercepts is convenient. Let x represent the average value of a particular extensive property *per mole of mixture*, i.e., $X = (n_1 + n_2)x$; then

$$\bar{X}_2 = \left(\frac{\partial X}{\partial n_2} \right)_{n_1} = x + (n_1 + n_2) \left(\frac{\partial x}{\partial n_2} \right)_{n_1}. \quad (172)$$

The mole fraction x_1 of the component (1) is equal to $n_1/(n_1 + n_2)$, and differentiation, keeping n_1 constant, gives

$$\begin{aligned} \partial x_1 &= -n_1 \partial n_2 / (n_1 + n_2)^2 = -x_1 \partial n_2 / (n_1 + n_2); \\ \therefore (n_1 + n_2) \left(\frac{\partial x}{\partial n_2} \right)_{n_1} &= -x_1 \left(\frac{\partial x}{\partial x_1} \right)_{n_1}, \end{aligned} \quad (173)$$

and hence from (172) it follows that

$$\bar{X}_2 = x - x_1 \left(\frac{\partial x}{\partial x_1} \right)_{n_1}. \quad (174)$$

The values of x , for various mixtures n_1 and n_2 , with the former constant, are plotted against the mole fraction x_1 , as shown by the curve MN in Fig. 39. Let

O be any point at which the partial molar quantity is to be determined; at O draw the tangent CD and the horizontal line EF parallel to the base line AB . The slope of CD is $(\partial x/\partial x_1)$, and so CE is equal to $x_1(\partial x/\partial x_1)$ at O . Since AE is the value of x at that point, it is obvious, from (174), that the distance AC gives the partial molar quantity \bar{X}_2 . In an exactly similar manner, it can be shown that BD is equal to \bar{X}_1 for the mixture represented by O .

This method enables \bar{X}_1 and \bar{X}_2 to be obtained in one operation, but if a procedure is used which gives only \bar{X}_2 , for example, \bar{X}_1 may be evaluated by making use of the Gibbs-Duhem equation (159). Writing this in the form

$$-d\bar{X}_1 = \frac{n_2}{n_1} d\bar{X}_2 \quad (159a)$$

and integrating between the limits \bar{X}_1 and \bar{X}'_1 for one component and \bar{X}_2 and \bar{X}'_2 for the other, it is seen that

$$\bar{X}_1 - \bar{X}'_1 = \int_{\bar{X}_2}^{\bar{X}'_2} \frac{n_2}{n_1} d\bar{X}_2. \quad (175)$$

If values of n_2/n_1 are plotted against the corresponding results for \bar{X}_2 , then the area under the curve between \bar{X}_2 and \bar{X}'_2 gives $\bar{X}_1 - \bar{X}'_1$. If the second state represents the pure component (1), then \bar{X}'_1 is equal to X_1^0 , and \bar{X}_1 may thus be determined.

Thermal Properties of Homogeneous Mixtures.—The partial thermal properties of homogeneous mixtures, particularly of liquid systems, i.e., solutions, are of special interest. If H is the heat content of a solution made up of n_1 moles of component (1) and n_2 moles of (2), and H_1^0 and H_2^0 are the molar heat contents of the pure substances, at the same temperature and pressure, then when the two components are mixed the increase of heat content ΔH is given by

$$\Delta H = H - n_1 H_1^0 - n_2 H_2^0. \quad (176)$$

This quantity is the **total** or **integral heat of solution**; it is the difference in the heat contents of the solution and of the components of which it is made up. The values of the integral heat of solution are usually recorded for 1 mole of the particular component under consideration, i.e., $\Delta H/n_1$ or $\Delta H/n_2$, respectively. It represents the total heat absorbed when 1 mole of the given component is mixed with the requisite amount of the other to form the prescribed solution.

If (176) is differentiated with respect to n_2 , while n_1 is maintained constant, as also are the temperature and pressure, the result is

$$\left(\frac{\partial(\Delta H)}{\partial n_2} \right)_{n_1} = \left(\frac{\partial H}{\partial n_2} \right)_{n_1} - H_2^0. \quad (177)$$

The left-hand side of this equation is the **partial** or **differential heat of solution** of component (2), which may be represented by the symbol \bar{Q}_2 , i.e.,

$$\bar{Q}_2 = \left(\frac{\partial(\Delta H)}{\partial n_2} \right)_{n_1} \quad (178)$$

Since $(\partial H/\partial n_2)_{n_1}$, at constant T and P , is the partial molar heat content \bar{H}_2 , it follows from (177) and (178) that

$$\bar{Q}_2 = \bar{H}_2 - H_2^0. \quad (179)$$

The differential heat of solution, as given by (178) or (179), is seen to be the heat absorbed when 1 mole of component (2) is added to a very large quantity of the solution at the specified concentration.

Since no distinction has been made between the two components, it is possible to derive the relationship, analogous to (179), viz.,

$$\bar{Q}_1 = \bar{H}_1 - H_1^0. \quad (180)$$

For many purposes it is convenient to regard one component, viz. (1), of a mixture as the solvent, whereas (2), present in lesser amount, may be considered to be the solute (cf. p. 626). In that case \bar{Q}_1 , i.e., the heat absorbed when 1 mole of solvent is added to a large bulk of a given solution, becomes the **partial** or **differential heat of dilution** at the particular concentration, which is assumed to remain constant.

When an appreciable amount of solvent is added to a solution, so that the concentration changes, the heat absorbed is the **total** or **integral heat of dilution**. By the first law, it must be equal to the difference in the heats of formation, i.e., the integral heats of solution, of the final and initial solutions, and if these are $\Delta H'$ and ΔH , respectively, then

$$\text{Integral heat of dilution} = \Delta H' - \Delta H. \quad (181)$$

Utilizing (155), it is seen that

$$\Delta H = n_1 \bar{H}_1 + n_2 \bar{H}_2 \quad \text{and} \quad \Delta H' = n_1' \bar{H}_1' + n_2 \bar{H}_2',$$

and hence, by (176), (179) and (180)

$$\Delta H' - \Delta H = (n_1' \bar{Q}_1' + n_2 \bar{Q}_2') - (n_1 \bar{Q}_1 + n_2 \bar{Q}_2), \quad (182)$$

where $\Delta H' - \Delta H$ is the heat absorbed when $n_1' - n_1$ moles of solvent are added to a solution containing originally n_1 moles of solvent and n_2 moles of solute.

The absolute partial heat contents of the components of a solution cannot be determined and so it is the practice to record the **relative partial molar heat content**; this is the difference between the value for a given solution and that in a reference state. It is convenient to choose an infinitely dilute solution for the purpose, so that the reference state of the solvent (1) is the pure liquid, and that of the solute (2) is infinite dilution (see p. 685). The relative partial molar heat content \bar{L}_1 of the solvent is thus

$$\bar{L}_1 = \bar{H}_1 - H_1^0 = \bar{Q}_1, \quad (183)$$

which by (180) is equal to the differential heat of dilution. For the solute \bar{L}_2 is given by

$$\bar{L}_2 = \bar{H}_2 - \bar{H}_2^0, \quad (184)$$

where \bar{H}_2^0 is the partial molar heat content of the solute (2) at infinite dilution. If H_2^0 , the heat content of (2) in the pure state, is added to and subtracted from (184), it is seen that

$$\bar{L}_2 = (\bar{H}_2 - H_2^0) - (\bar{H}_2^0 - H_2^0) = \bar{Q}_2 - \bar{Q}_2^0, \quad (185)$$

where \bar{Q}_2^0 is the differential heat of solution at infinite dilution. By equation (159), $n_1 d\bar{H}_1 + n_2 d\bar{H}_2 = 0$, and hence from (183) and (184), noting that H_1^0 and \bar{H}_2^0 are constants, it follows that $n_1 d\bar{L}_1 + n_2 d\bar{L}_2 = 0$, so that \bar{L}_1 can be determined if \bar{L}_2 is known over a range of concentrations, or *vice versa*.

By (16), the heat capacity C_P of a solution containing n_1 moles of solvent and n_2 of solute is $\partial H/\partial T$; if this is differentiated with respect to n_1 , keeping n_2 constant, then at constant T and P ,

$$\frac{\partial^2 H}{\partial T \partial n_1} = \left(\frac{\partial C_P}{\partial n_1} \right)_{n_2} = (\bar{C}_P)_1, \quad (186)$$

where $(\bar{C}_P)_1$ is the partial molar heat capacity of the solvent. By definition, the partial molar heat content of the latter, i.e., \bar{H}_1 is $(\partial H/\partial n_1)_{n_2}$, and differentiation with respect to T gives

$$\frac{\partial^2 H}{\partial T \partial n_1} = \frac{\partial \bar{H}_1}{\partial T}, \quad (187)$$

which must, by (186), be equal to $(C_P)_1$. If now (183) is introduced it is seen that

$$\frac{\partial \bar{L}_1}{\partial T} = \frac{\partial \bar{H}_1}{\partial T} - \frac{\partial H_1^0}{\partial T} = (\bar{C}_P)_1 - (C_P^0)_1, \quad (188)$$

where $(C_P^0)_1$ is the molar heat capacity of the pure solvent. By a similar method it is possible to deduce, from (184), that

$$\frac{\partial \bar{L}_2}{\partial T} = \frac{\partial \bar{H}_2}{\partial T} - \frac{\partial \bar{H}_2^0}{\partial T} = (\bar{C}_P)_2 - (\bar{C}_P^0)_2, \quad (189)$$

where $(\bar{C}_P)_2$ and $(\bar{C}_P^0)_2$ are the partial molar heat capacities of the solute in the given solution and at infinite dilution respectively; the difference is called the **relative partial molar heat capacity**.

REFERENCES

- Bichowsky, F. R. and Rossini, F. D. "The Thermochemistry of Chemical Substances."
 Brønsted, J. N. "Physical Chemistry." (Trans. by R. P. Bell.)
 Butler, J. A. V. "The Fundamentals of Chemical Thermodynamics," Vols. I and II.
 Guggenheim, E. A. "Modern Thermodynamics."
 Lewis, G. N. and Randall, M. "Thermodynamics, etc."
 MacDougall, F. H. "Thermodynamics and Chemistry"; "Physical Chemistry."
 Partington, J. R. "Chemical Thermodynamics."
 Rossini, F. D. *Chem. Rev.* 18, 233 (1936); 27, 1 (1940).
 Steiner, L. E. "Introduction to Chemical Thermodynamics."
 Taylor, H. S. and Glasstone, S. "Treatise on Physical Chemistry," 3rd. ed., Vol. I, Chap. III (H. S. Taylor.)
 Thomsen, J. "Thermochemistry." (Trans. by K. A. Burke.)
 White, W. P. "The Modern Calorimeter."

CHAPTER IV

THE GASEOUS STATE

THE KINETIC THEORY OF GASES

States of Matter.—Matter can be divided roughly into three main categories, namely gases, liquids and solids; these are often described as the three chief states of matter. It is true that certain substances, such as vitreous materials and others, do not fall definitely into one or other of these classes, but it is convenient, nevertheless, to consider the properties of matter under the three main headings. The gaseous state is characterized by a marked sensitivity of the volume to changes of temperature and pressure, and also by the fact that a gas has normally no bounding surface and so it tends to fill completely any available space. A liquid, like a gas, has no definite shape and, in fact, takes the shape of the vessel wherein it is placed; it has, however, a surface which places a limit on the extent of space it is able to occupy. It is this surface, as will be seen later, that is responsible for many of the characteristic properties of liquids. Solids differ markedly from liquids and gases because they have a definite shape; like liquids, however, their volume does not change appreciably as the result of changes in temperature and pressure. It will be evident that, on the whole, the properties of liquids are intermediate between those of gases and solids, and the natural order for treating the states of matter would be that of increasing complexity of structure, that is gas, liquid, solid. The unconventional method will be adopted here, however, of discussing first gases, then solids and finally liquids, for the reason that recent studies have shown that the structure of liquids can be best understood in the light of the known structure of solids.

Deviations from Boyle's Law.—According to Boyle's law (p. 191) the volume of a given mass of gas should be inversely proportional to the pressure, but the measurements on which the conclusion was based were of limited scope and accuracy. During the nineteenth century a number of workers, particularly H. V. Regnault (1847–1862) and E. H. Amagat (1880–1893), studied the compressibilities of various gases with great care, and it became evident that for real gases Boyle's law could only be regarded as a rough approximation. The magnitude and nature of the and volume, that is, PV , at constant temperature for various pressures; if Boyle's law were obeyed the values would be constant. The data for hydrogen and nitrogen at 0° c., and for carbon dioxide at 40° c., are given in Table 28, and plotted in Fig. 40, the value of PV at 1 atm. pressure being taken as unity in each case; the temperature 40° is chosen for carbon

TABLE 28. RELATIVE VALUES OF THE PRESSURE-VOLUME PRODUCT

Pressure	Hydrogen	Nitrogen	Carbon Dioxide
1 atm	1.0000	1.0000	1.0000
50	1.0330	0.9846	0.7413
100	1.0639	0.9846	0.2695
200	1.1336	1.0365	0.4087
400	1.2775	1.2557	0.7178
800	1.5665	1.7959	1.2990
1000	1.7107	2.0641	1.5525

dioxide as liquefaction occurs below 31°C . (see p. 429). The results in this table show that the departure from Boyle's law, especially at high pressures, is very considerable, and at 1000 atm. the product PV is nearly twice the value to be expected if the law were applicable. At the lower pressures, however, the deviations are not great; for example, within the pressure range of 1 to 10 atm. the deviations from Boyle's law are about 5 per cent, or less, for most gases, although the more easily liquefiable the gas the greater is the error.

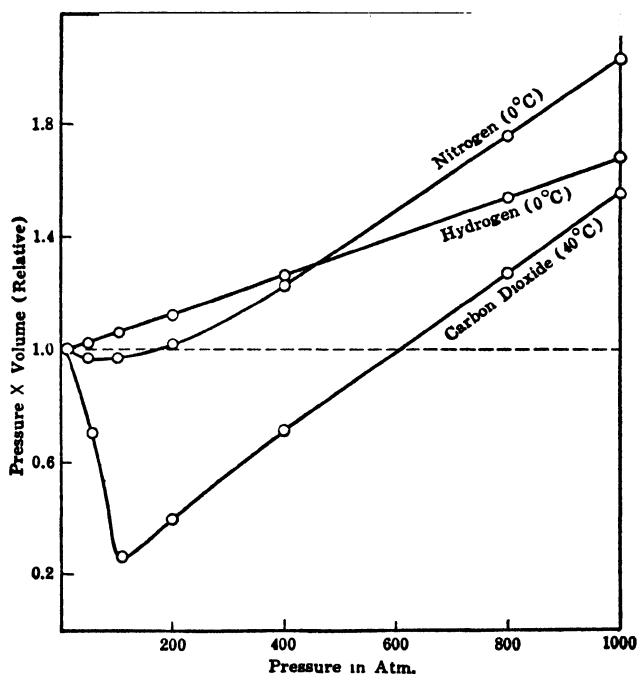


FIG. 40. Variation of pressure-volume product with pressure

An examination of Fig. 40 reveals the fact that whereas the value of PV for hydrogen increases continuously as the pressure is increased, with the other gases there is at first a decrease; this was discovered by

Regnault, who described hydrogen as "*un gaz plus que parfait*," because it was less compressible than the perfect, or ideal, gas obeying Boyle's law. It is now known that the behavior of hydrogen in this respect is not exceptional, for the type of variation of PV with pressure is determined by the temperature. Helium and neon are similar to hydrogen in being less compressible than an ideal gas at ordinary temperatures, but as the temperature is lowered the shape of the PV curve changes, and in each case it eventually becomes like that for nitrogen at 0° or carbon dioxide at 40° C. On the other hand, at sufficiently high temperatures the pressure-volume product of these latter, and other, gases increases right from the commencement as the pressure is raised. The curves in Fig. 41 show

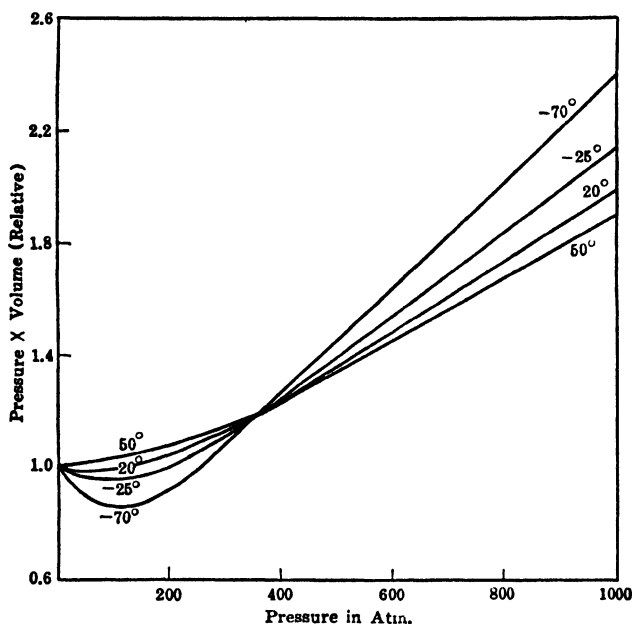


Fig. 41. Variation of pressure-volume product of nitrogen.

the values of PV , relative to those for 1 atm. in each case, for nitrogen gas at a series of temperatures; the change in the shape of the curves with increasing temperature is clearly seen. Analogous curves have been obtained for other gases, and it is probable that the behavior is universal.

The Boyle Temperature.—It is seen from Fig. 41 that as the temperature is raised the dip in the PV curve becomes smaller, and the position of the minimum moves to the left. At a certain temperature, when the minimum falls on the PV -axis, the curve must remain horizontal, that is, PV is constant, over an appreciable range of pressures. Within this range, therefore, Boyle's law is obeyed, and hence the temperature is called the **Boyle point**. Below the Boyle temperature the value of PV at first

decreases as the pressure is increased, but above it PV increases continuously. It is evident that hydrogen and helium have very low Boyle points, the values are approximately -165° and -240° c., respectively, but for nitrogen and other gases they are higher. In general, the more easily liquefiable gases have high Boyle points, whereas for those difficult to liquefy, e.g., hydrogen and helium, the Boyle temperatures are very low. The relationship between this temperature and other properties of gases will be considered later (p. 291).

Quite apart from the particular type of behavior shown by a gas, it is evident that the extent of the deviation from Boyle's law is small at low pressures, especially if the temperature is relatively high. It appears that gases which are not easily liquefied, such as hydrogen, helium, nitrogen, oxygen, carbon monoxide and nitric oxide, obey the law, within the limits of experimental error, at pressures less than about 0.1 atm. at ordinary temperatures. It is probable, therefore, that the lower the pressure the more closely do gases approach the ideal behavior required by Boyle's law, and corrections for the departure from the ideal are generally made by extrapolation to zero pressure (cf. p. 309).

Deviations from Gay-Lussac's Law.—According to Gay-Lussac's law (p. 191) the coefficients of volume expansion of all gases should be the same and independent of the pressure. The work of Amagat, as well as of more recent investigators, has shown that this is not true and that appreciable deviations are apparent at high pressures. In Table 29 are recorded the mean coefficients of expansion from 0° to 100° c. of hydrogen, nitrogen and carbon dioxide for pressures from 1 to 1000 atm. As with Boyle's law, the most easily liquefiable gases deviate most from ideal behavior. The results show that although the coefficients of expansion differ

TABLE 29. MEAN COEFFICIENTS OF EXPANSION FROM 0° TO 100° C.

Pressure	Hydrogen	Nitrogen	Carbon Dioxide
1 atm.	366×10^{-6}	368×10^{-6}	372×10^{-6}
50	358	400*	811*
200	314	433	1115
500	278	315	349
1000	218	200*	206

* Estimated values.

appreciably at high pressures, the values at low pressures all approach the same value; by extrapolation to zero pressure the coefficient of expansion of an ideal gas, as recorded in p. 192, has been estimated as 0.0036609.

Deviations from Avogadro's Law.—The accuracy of Avogadro's law may be tested by measuring the volume occupied by 1 mole of a gas at a temperature of 0° c. and 1 atm. pressure; the molecular weight is determined from the sum of the atomic weights of the constituent atoms. If the law were strictly true, the volume would be independent of the nature of the gas; that this is only approximately so is shown by the data in

TABLE 30. THE APPLICABILITY OF AVOGADRO'S LAW

Gas	Molecular weight	Wt. of 1 liter at 0° c. and 1 atm.	Vol. of 1 mole at 0° c. and 1 atm.
Hydrogen	2.0160	0.08988	22.425
Nitrogen	28.016	1.2506	22.402
Oxygen	32.000	1.4290	22.394
Carbon dioxide	44.010	1.9767	22.264
Ammonia	17.032	0.7712	22.084
Methyl chloride	50.491	2.3076	21.879

Table 30. The largest deviations, as before, are shown by the more easily liquefiable gases. The figures in the last column are seen to be only approximately constant, so that Avogadro's law is not exactly obeyed by real gases at ordinary temperatures and pressures; at low pressures, however, the deviations become less. By making due allowance for the departure from ideal behavior, the agreement with Avogadro's law is so good that accurate molecular weights have been evaluated from density measurements (p. 309) for a number of gases.

The Kinetic Theory of Gases.—The properties of an ideal gas, and the reasons for the departure of real gases from ideal behavior, are best interpreted by means of the **kinetic theory of gases**: this theory has developed from the ideas chiefly of D. Bernouilli (1738), J. J. Waterston (1845; published 1892), K. A. Krönig (1856), and R. Clausius (1857), but it was given precise mathematical form by J. Clerk Maxwell (1860) and L. Boltzmann (1868). It is supposed that a gas consists of a large number of very small, *perfectly elastic* particles, which may be identified with the chemical molecules, moving about in all directions. For an ideal gas, the theory postulates that the molecules are so small that their actual volume is negligible in comparison with the total volume of the gas, and that they exert no attraction upon one another. As a result of their continual movement in all directions, the molecules will frequently collide with each other and with the walls of the containing vessel, and it is the latter elastic impacts which, according to the kinetic theory, are responsible for the pressure exerted by a gas. Increase of temperature will result in more vigorous movement of the particles, so that if the volume remains constant the pressure of the gas must increase; further, if the molecules are made to occupy a smaller volume they will hit the walls of the vessel more frequently and so an increase of pressure will result.

Random Movement of Molecules.—If the molecules of a gas are in motion, then it can be readily seen from general considerations, although it may also be proved more rigidly, that they cannot all have the same velocity, neither will they tend to move in any one direction in preference to another. Suppose the molecules in a given vessel all moved in parallel lines with the same speed; the resulting system would be unstable, for the slightest disturbance, sufficient to cause a single molecule to deviate from its path, would bring about chaotic movement. The deviated molecule would strike another, and as a result of

the impact both would move off in different directions with different velocities. These two molecules would then undergo impacts with others, with the result that in a short time the movement of all the molecules in the given vessel would be completely random. By the principle of the conservation of momentum, as applied to collisions between elastic bodies, it can be shown that even if all the molecules in a gas had the same initial velocity, then as a result of impacts this condition would be destroyed. A simple illustration is the following: consider two molecules, *A* and *B*, having the same mass and moving with the same speed *c* in directions at right angles to each other. At the moment of impact the line of centers corresponds to the direction of motion of *A*. As a consequence of the collision the molecule *A* will be brought to rest, whereas *B* will move off in a direction making an angle of 45° with the original line of centers, and its velocity will be increased to $c\sqrt{2}$. As a result of further collisions with other molecules in the gas, *A* will acquire momentum and commence to move again, whereas *B* may either have its speed increased further or decreased, depending on the nature of its subsequent impacts. If a large number of molecules are considered, it is evident that the net result will be movement in all directions, and also a non-uniformity of velocity, for the latter may vary from zero upward. Actually only a small proportion of the molecules have very low or very high speeds, the majority having velocities in the vicinity of the average value. The question of the actual distribution of velocities will be considered later (p. 264).

Velocity Components.—The velocity *c* of a given molecule may be represented by a vector having components *u*, *v* and *w* along three axes perpendicular to one another, so that

$$c^2 = u^2 + v^2 + w^2, \quad (1)$$

and the same applies to all molecules having the velocity *c*, no matter what the direction, although the separate components, *u*, *v* and *w*, may be different. Since all directions are equally possible, the *mean* value of the component parallel to any axis, e.g., \bar{u} ,* must be zero, since there are equal probabilities for the components in opposite directions; that is $+u$ and $-u$ are equally probable. On the other hand, $\bar{u^2}$, which represents the mean of the squares of the velocity components parallel to one axis, is not zero, since the squares are always positive; further, the values are the same for the other two axes, for no direction of motion is preferred to any other. It follows, therefore, that

$$\bar{u^2} = \bar{v^2} = \bar{w^2}, \quad (2)$$

and hence by (1) each of these is equal to $c^2/3$, all the molecules being assumed to have the same velocity *c*.

The Pressure of a Gas.—By considering the change of momentum resulting from the impacts of the molecules on the walls of the containing vessel, it is possible to calculate theoretically the pressure exerted on the

* A bar over the top of a letter is the conventional method of representing a mean value.

basis of the assumptions of the kinetic theory of gases. Several methods have been proposed; the most rigid are somewhat complicated whereas the simple treatment found in elementary text-books, although giving the correct pressure, is not altogether satisfactory especially as it leads to a wrong estimate of the number of collisions with the walls of the vessel. This objection does not apply to the deduction given below. Suppose a given volume, V cc., of gas contains N molecules of which the number n_a per cubic centimeter have the same velocity c_a irrespective of direction; of these let n_1 molecules per cc. have a component u_1 in a given direction, let n_2 have the component u_2 in the same direction, and so on. Consider

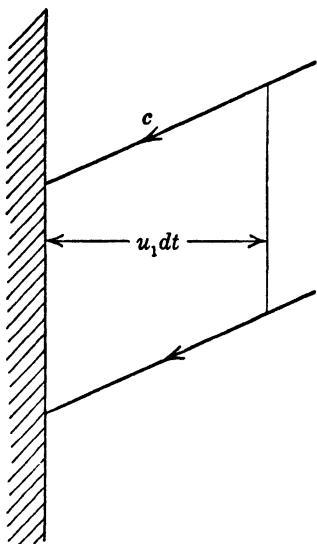


FIG. 42. Impact of molecules on a wall

the impact on 1 sq. cm. of a wall of the containing vessel at right angles to the direction of the n_1 molecules with the velocity component u_1 perpendicular to this wall (Fig. 42). In a small time dt the number of molecules striking the 1 sq. cm. of wall will be all those which at the beginning of the time interval were contained in the volume element depicted; the length of this element is $u_1 dt$, and since its cross sectional area is 1 sq. cm., this will also be its volume. The number of molecules, of the kind considered, in the volume element will then be $n_1 u_1 dt$, and since this number strike the wall in time dt , the 1 sq. cm. of wall will be struck at the rate of $n_1 u_1$ molecules in unit time.

The molecules are supposed to consist of perfectly elastic particles, and so the speed will be the same before and after impact with the wall, and the angle between the given axis and the original and final directions will be the same. The momentum of a molecule of mass m will therefore be mu_1 parallel to the axis and towards the wall before impact, and mu_1 in the opposite direction after impact; the change of momentum perpendicular to the wall is consequently $2mu_1$ for each impact. The change of momentum in unit time, that is for $n_1 u_1$ impacts, is thus $2mn_1 u_1^2$ for all the molecules having the same velocity component u_1 in the given direction. Since all directions of incidence on the wall, over a hemisphere, are possible, it follows that

$$\text{Change of momentum in unit time} = 2m \sum n u^2. \quad (3)$$

The summation term can be defined by

$$\sum n u^2 = \frac{1}{2} (n_1 u_1^2 + n_2 u_2^2 + n_3 u_3^2 + \dots), \quad (4)$$

where the factor $\frac{1}{2}$ is introduced because the summation occurs over a

hemisphere, i.e., only for molecules moving *towards* the 1 sq. cm. of wall. The mean square of the velocity component $\overline{u_a^2}$ is given by

$$\overline{u_a^2} = \frac{n_1 u_1^2 + n_2 u_2^2 + n_3 u_3^2 + \dots}{n_1 + n_2 + n_3 + \dots}, \quad (5)$$

and since $n_1 + n_2 + n_3 + \dots$, that is the sum of all the molecules having the same velocity c_a , is n_a , as defined above, it follows that (3) may be written:

$$\text{Change of momentum in unit time} = 2m \times \frac{1}{2} n_a \overline{u_a^2} = m n_a \overline{u_a^2}. \quad (6)$$

It has been already seen that for a given velocity c , the value of $\overline{u^2}$ is equal to $c^2/3$; hence it follows that

$$\text{Change of momentum in unit time} = \frac{1}{3} m n_a c_a^2. \quad (7)$$

In order to determine the change of momentum resulting from the impact of all the molecules, it is necessary to sum the terms in (7) for all possible velocities, and so

$$\text{Total change of momentum in unit time} = \frac{1}{3} m n \overline{c^2}, \quad (8)$$

where n is the total number of molecules per unit volume, and $\overline{c^2}$ is the mean square velocity for all the molecules in the given gas, defined in a manner analogous to (5). Since the total volume of gas is V and this contains N molecules, n is equal to N/V , and so (8) becomes

$$\text{Total change of momentum in unit time} = \frac{1}{3} m N \overline{c^2} / V. \quad (9)$$

It is known from mechanics that the force acting on a body is equal to the rate of change of momentum, i.e., the change of momentum in unit time; further, pressure is the force acting on unit area, and so pressure is equal to the change of momentum in unit time per unit area. This is actually the quantity given by (9), since it represents the change of momentum resulting from impacts on 1 sq. cm. of the wall; hence it follows that

$$P = \frac{1}{3} m N \overline{c^2} / V \quad (10)$$

or

$$PV = \frac{1}{3} m N \overline{c^2} \quad (11)$$

$$= \frac{2}{3} \cdot \frac{1}{2} m N \overline{c^2}. \quad (12)$$

The quantity $\frac{1}{2} m N \overline{c^2}$ is equal to the kinetic energy of all the molecules present in a given volume, the mean energy per molecule being $\frac{1}{2} m \overline{c^2}$, and so the product of pressure and volume of the gas is equal to two-thirds of the total kinetic energy.

Boyle's Law.—In the simple ideal type of gas considered by the kinetic theory, the energy of the molecules is due to their translation only, and hence it is kinetic energy. If the gas is heated the temperature rises

and the thermal energy must presumably be converted into kinetic energy which increases correspondingly. It is reasonable to suppose, therefore, as a consequence of the kinetic theory, that the temperature of a gas and the kinetic energy of its molecules are directly related, and provided the temperature of a given quantity of gas remains constant, so also does the total kinetic energy. It follows, therefore, that at a definite temperature the right-hand side of (12) is constant, and hence the value of PV should be constant; this is the mathematical statement of Boyle's law, which has thus been deduced from the kinetic theory of gases.

Avogadro's Law.—If two different gases having molecules of the *same weight* are enclosed in a vessel, the average kinetic energies per molecule, even if different at the commencement, must as a result of collisions eventually become the same. Since the interchange of energy in elastic collisions is dependent on the mass and speed of the colliding particles, but is independent of their nature, there is no kinetic distinction between the molecules in the mixture, and the mean kinetic energy of the molecules of one gas is the same as for the other. At the same temperature, therefore, it may be supposed that the mean kinetic energy per molecule, $\frac{1}{2}m\bar{c}^2$, is constant for the two gases having molecules of the same weight. It was shown by J. C. Maxwell that this important rule is applicable to all molecules, irrespective of their masses; the actual proof is based on the law of distribution of velocities (see p. 264), but it is too complicated to be given here. For any two gases, indicated by the suffixes 1 and 2, it is possible to write

$$P_1V_1 = \frac{1}{3}N_1m_1\bar{c}_1^2 \quad \text{and} \quad P_2V_2 = \frac{1}{3}N_2m_2\bar{c}_2^2, \quad (13)$$

and if the temperature is the same, Maxwell's treatment requires $\frac{1}{2}m_1\bar{c}_1^2$ to be equal to $\frac{1}{2}m_2\bar{c}_2^2$, so that if both gases are at the same pressure, i.e., $P_1 = P_2$, and occupy the same volume, $V_1 = V_2$, it follows that N_1 and N_2 must be equal. In other words, it is seen that for two gases at the same temperature and pressure, equal volumes should contain equal numbers of molecules; this is, of course, Avogadro's law (p. 193).

Kinetic Energy and Temperature.—A direct connection between the kinetic energy or speed of the molecules of a gas and its temperature may be obtained by introducing the thermodynamic definition of an ideal gas as one whose internal energy is independent of its volume, at constant temperature. As mentioned on p. 236, this postulate, together with Boyle's law, includes Gay-Lussac's law, and a combination of these two laws with that of Avogadro leads to the familiar equation

$$PV = RT \quad (14)$$

for 1 mole of an ideal gas. From the kinetic theory, it has been seen [equation (12)] that

$$PV = \frac{2}{3} \cdot \frac{1}{2}Nmc^2,$$

where N is now the number of molecules in 1 mole of gas, i.e., the Avogadro number; hence,

$$\frac{2}{3} \cdot \frac{1}{2} N m \bar{c}^2 = RT. \quad (15)$$

If the total kinetic energy of the molecules, that is, $\frac{1}{2} N m \bar{c}^2$, is represented by E_K , it follows that

$$E_K = \frac{3}{2} RT, \quad (16)$$

so that the kinetic energy of 1 mole of any gas is directly proportional to its absolute temperature; since N and $\frac{1}{2} m \bar{c}^2$ are the same for all gases, (16) is independent of the nature of the gas. As the kinetic energy involves the square of the velocity, it is clear that the molecular velocity itself will be proportional to the square root of the absolute temperature. The proportionality constant, however, depends on the molecular weight, as will be seen below.

Molecular Velocity.—By writing (12) in the form

$$\bar{c}^2 = 3PV/mN,$$

it follows that

$$\sqrt{\bar{c}^2} = \sqrt{3PV/mN} = \sqrt{3PV/M}, \quad (17)$$

where M , which is equal to mN , is the mass of gas in volume V ; if the latter is the gram molecular volume, then M is the molecular weight. In that case PV may be replaced by RT as indicated above, so that

$$\sqrt{\bar{c}^2} = \sqrt{3RT/M}. \quad (18)$$

This equation not only shows the relationship between velocity, temperature and molecular weight, but it allows the square root of the mean square velocity of the molecules of any gas to be calculated.* For molecular oxygen, for example, M is 32 and, assuming the temperature to be 0°C. , it follows that

$$\begin{aligned} \sqrt{\bar{c}^2} &= (3 \times 8.31 \times 10^7 \times 273/32)^{\frac{1}{2}} \\ &= 4.61 \times 10^4 \text{ cm. per sec. at } 0^\circ. \end{aligned}$$

The value of R is, of course, in ergs, as given on p. 194, so that the velocity should be in c.g.s. units, i.e., cm. per sec. If the temperature is increased it is seen from (18) that the speed of the molecules will increase, but the value is independent of the pressure of the gas. At first sight this appears to be contrary to (17), but further examination shows that it involves PV which remains constant in spite of changes of pressure.

* It should be noted that the square root of the mean square velocity is not equal to the mean velocity, \bar{c} ; the latter, as will be seen later, is $\sqrt{8/3\pi}$ times the former for an actual gas. The difference between the two quantities may be readily seen by considering three molecules having velocities 1, 2 and 3, respectively; \bar{c}^2 is $(1^2 + 2^2 + 3^2)/3 = 4.66$, and its square root is 2.16, whereas the mean velocity \bar{c} is 2.0.

Experimental Determination of Molecular Velocity.—The first successful attempt to measure molecular velocities was made by O. Stern (1920): the apparatus consisted of a framework, shown in plan in Fig. 43, containing two pairs of slits S_1 and S_2 . At the center was a platinum filament F coated with silver, so that on heating electrically silver atoms evaporated off, and those that succeeded in passing through the slits were deposited on the plates PP ; the path of such atoms would be along the line PPF . The apparatus was enclosed in an evacuated vessel, so that the number of collisions between silver atoms was negligible. The framework was then rotated rapidly about a vertical axis through F , as shown by the arrows. The direction of movement of the silver atoms is unaffected, but in the time taken for them to travel from F to P , the latter has moved as a result of the rotation of the framework, and the apparent path of the atoms is shown by the broken line. If the distance between the positions of the silver deposits with the apparatus stationary and when rotating is measured, and the speed of rotation is known, the time for the plate P to move this distance, and hence for the silver atoms to travel from F to

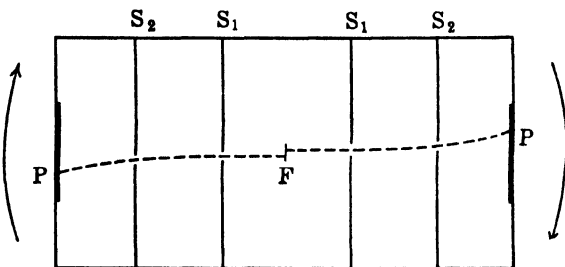


FIG. 43. Determination of molecular velocities (Stern)

P , is known. The velocity of the atoms can then be calculated. The atoms do not all have the same speed and so the silver deposit is actually spread out in a narrow band, instead of being a fine line. The most intense portion is produced by atoms having approximately the average or, more exactly, the most probable, velocity (see p. 267). Most of the silver atoms were found to have a velocity between 560 and 640 meters per sec. The temperature of the wire was *estimated* to be about 1200°C ., at which the speed by the kinetic theory equation should be 584 meters per sec. Other methods available for the determination of molecular velocities are described on p. 269.¹

Diffusion of Gases.—By utilizing the fact that M/V represents the density (ρ) of the gas, (17) becomes

$$\sqrt{\bar{c^2}} = \sqrt{3P/\rho}, \quad (19)$$

so that the speed of the molecules is seen to be inversely proportional to the square root of the gas density. Since it is highly probable that the diffusion of a gas is related to the speed of its molecules, it is apparent that the kinetic theory provides a theoretical basis for T. Graham's **law of diffusion** (1829), according to which *the rate of diffusion of a gas is inversely proportional to the square root of its density*; alternatively, it fol-

lows from (18) that the rate of diffusion should be inversely proportional to the square root of the molecular weight. Objection has been made to the kinetic theory because the speeds of molecules are very many times greater than the rate at which the gas as a whole is able to diffuse. It must be borne in mind, however, that a gas molecule does not move very far in a straight line before it collides with another. These collisions are so frequent, as will be seen shortly, that the whole gas may move forward at a relatively slow rate in spite of the high velocities of the individual molecules. By making use of the dependence of diffusion on density, the separation of two gases is possible; the process is called **atmolysis**, and it has been applied with success to the separation of isotopes (p. 153) and other gaseous mixtures.

Brownian Movement and the Avogadro Number.—One of the most striking and important pieces of evidence in favor of the kinetic theory and of the real existence of molecules is based on the determination of the Avogadro number and the study of the movement of small particles. In 1827, the botanist Robert Brown noticed that pollen grains when suspended in water exhibited continual and haphazard motion in all directions; in fact their behavior was, on a large scale, exactly what one would expect from a molecule if the kinetic theory were correct. This erratic motion, now called the **Brownian movement**, has been observed with small suspended particles of all kinds, and is quite independent of the nature of the substance and of the medium in which it is suspended, whether liquid or gaseous. As might be anticipated, increasing the viscosity of the medium results in a slowing down of the Brownian movement. The explanation of the phenomenon, first suggested by C. Wiener (1863), and later independently by W. Ramsay (1879) and by G. Gouy (1888), is that the movement is caused by the continual bombardment of the suspended particles by the molecules of the medium in which they are suspended. That is to say, the continual and erratic motion of the particles is a direct consequence of the continual and erratic movements of the molecules surrounding them. The molecules are too small to be visible but the influence of their impacts is made evident by the movement of the relatively large particles seen with the aid of the microscope.

In order to provide evidence for this view, and incidentally for the kinetic theory generally, which it must be pointed out is assumed here to apply to liquids as well as gases, J. Perrin (1908) undertook a series of investigations of various aspects of the Brownian movement. Consider a vertical cylinder of uniform cross section containing a gas in which the molecules have come into equilibrium with gravity, the only external force acting on them; let P be the gas pressure at the height h from the bottom of the cylinder, and let $P - dP$ be the pressure at the height $h + dh$. It should be noted that, as with the vertical distribution of the pressure of the atmosphere, the higher the level the lower the gas pressure. Since equilibrium exists in the cylinder, it is evident that the difference in pressure of the two layers dh apart, acting upward, must be balanced by the gravitational force, acting downward, of the molecules contained in this layer. If the area of cross section of the cylinder is a , and the density of the gas is ρ at the pressure P , then the force upward is adP , and downward it is $apgdh$, where $apgdh$ is the mass of the gas, and g is the acceleration

due to gravity; hence

$$-adP^* = \dot{a}\rho gdh; \quad (20)$$

$$\therefore -dP = \rho gdh. \quad (21)$$

If M is the molecular weight of the gas and V the volume of 1 mole, then ρ is equal to M/V ; further, it has been shown by the kinetic theory that for an ideal gas PV is equal to $\frac{2}{3}E_K$ where E_K is the kinetic energy of all the gas molecules in 1 mole (p. 251). It follows, therefore, that ρ is equal to $3MP/2E_K$, and substitution in (21) gives

$$-dP = \frac{3MPg}{2E_K} dh \quad (22)$$

or

$$-\frac{dP}{P} = \frac{3Mg}{2E_K} dh. \quad (23)$$

If P_0 is the pressure at any arbitrary zero level, where h is 0, and P the pressure at the height h , then it follows on integration that

$$\ln \frac{P_0}{P} = \frac{3Mgh}{2E_K}. \quad (24)$$

Since the pressure is proportional to n , the number of molecules per unit volume [cf. equation (8)], it follows that

$$\ln \frac{n_0}{n} = \frac{3Mgh}{2E_K}, \quad (25)$$

where n_0 refers to the arbitrary zero level and n to the height h . In general the same equation will apply when n_0 and n represent the numbers of molecules per unit volume at any two levels at a distance h apart.

This equation, which resembles the familiar hypsometric formula, describes the vertical distribution of molecules under the influence of gravity, and Perrin supposed that it would also represent the distribution of microscopic particles when suspended in a liquid. This assumption is not unreasonable, as an analogous equation could be deduced by applying the kinetic treatment to microscopic particles considered as elastic spheres which have no appreciable volume and do not attract one another. For suspensions of this kind, the quantity Mg in (25), representing the gravitational force on all the N molecules in 1 mole, would have to be replaced by $Nmg(1 - \rho'/\rho)$ where N is the Avogadro number, m is the mass of a single particle, and ρ and ρ' are the densities of the particles and liquid media in which they are suspended, respectively. The introduction of ρ'/ρ is necessary, in accordance with the principle of Archimedes, because of the upward thrust on the particles due to the displacement of their own volume of liquid. It is now necessary to consider E_K , the kinetic energy of the N particles. According to Maxwell the mean kinetic energy of any molecule, that is $\frac{1}{2}m\bar{c}^2$, is the same for all gases at a given temperature (see p. 252). This rule of the constancy of the kinetic energy does not stipulate any limit to the mass of the molecule, and so it should apply equally to very large and very small molecules. Following out this argument, Perrin assumed that the kinetic energy of Brownian movement of a microscopic particle would be the same as that of a molecule. This assumption is in harmony with the view that the

* The sign must be included to show that the pressure decreases as the height increases.

Brownian movement is brought about by the impacts of the molecules of the medium in which the particle is suspended; the particle then behaves as a very large molecule, and its kinetic energy should be the same as the average for the surrounding molecules. If this is the case, then for the vertical distribution of a suspension the value of E_K in (25) is the same as the kinetic energy of N molecules, that is, of all the molecules in 1 mole of gas, and this, as already seen [equation (16)], is equal to $\frac{3}{2}RT$. Making the necessary alterations in (25), therefore, the equation for the numbers of particles n_0 and n at levels separated by a vertical distance h is given by

$$\ln \frac{n_0}{n} = \frac{Nmg h}{RT} \left(1 - \frac{\rho'}{\rho} \right). \quad (26)$$

If r is the radius of a particle, then its volume would be $\frac{4}{3}\pi r^3$ if it were spherical, and its mass m would be $\frac{4}{3}\pi r^3 \rho$, since ρ is its density; substituting this value in (26) and rearranging, it follows that

$$\frac{RT}{N} \ln \frac{n_0}{n} = \frac{4}{3}\pi r^3 g h (\rho - \rho'). \quad (27)$$

Examination of (27) shows that if it were possible to determine the ratio of the number of particles, of known radius and density, at two different levels of a suspension in equilibrium with gravity, a method would be available for the determination of the Avogadro number. Success in this connection was first achieved by J. Perrin and his collaborators. The suspensions* used consisted of gamboge and mastic, the particles being obtained of a uniform size by a process of fractional centrifuging. Three methods were used to determine the densities of the particles. The first was the same as that employed for insoluble powders; the densities of the water (d_0) and of the suspension in water (d_1) were first determined. A known weight w of the suspension was then evaporated to dryness and heated at 110° to constant weight (w_p); this gave the amount of solid present, and hence the weight of water ($w - w_p$) could be obtained by subtraction. Knowing the densities of water and of suspension and the composition of the latter, the density of the particles (ρ) could be determined by assuming that there is no net volume change when the suspension is made up; thus,

$$\frac{w - w_p}{d_0} + \frac{w_p}{\rho} = \frac{w}{d_1} \quad (28)$$

volume of
volume of
volume of
water
particles
suspension

All the data are thus available for the evaluation of ρ . In the second method, the glassy solid obtained on evaporating the suspension to dryness was broken up into small pieces and placed in water; sufficient potassium bromide was then added so that the fragments neither rose nor sank. The density of the solid was then the same as that of the potassium bromide solution, which could be determined in the ordinary way. If the suspended granules have the same density as the solid obtained on evaporation, the required density is that of the bromide solution. The third method involved the addition of the potassium bromide to the suspension itself until the particles exhibited no movement

* The suspensions employed by Perrin are often described as "emulsions," but this description is probably incorrect for most of the substances studied.

when centrifuged; the particles were then of the same density as the solution in which they were suspended. The concordance of the results is shown by the fact that for gamboge granules the three methods gave the values 1.1942, 1.194 and 1.195 respectively.

The radius of the particles was also measured in three ways. First, a small amount of dilute suspension was allowed to evaporate on a microscope slide; when nearly dry the granules ran together and collected in rows. The size of each particle could then be obtained by counting, under the microscope, the number in a row of measured length, or included in a definite area. In the second method, Perrin utilized the fact that in a 0.01 *N* acid medium the particles adhere to the walls of the glass vessel containing the suspension when they strike them; it is thus possible to count the total number of granules in a given amount of suspension. Knowing the volume of the solid contained therein, as found above, the volume of each particle could be ascertained. Perrin also made use of Stokes's law, the reliability of which he confirmed, for determining the size of suspended particles. If a suspension is shaken, so that the particles are uniformly distributed, then on standing the upper layers will clear gradually as the equilibrium distribution under the influence of gravity is attained. If u is the uniform rate at which the particles fall during the early stages, then according to Stokes's law the force acting, which results in their motion, is equal to $6\pi\eta ru$ where η is the viscosity of the medium and r the radius of the granules. This force must be equal to the effective gravitational pull downward on the particle, that is,

$$6\pi\eta ru = \frac{4}{3}\pi r^3(\rho - \rho')g, \quad (29)$$

and so if u is measured by observing the clearing of the upper layers, and the viscosity of the medium is known, the radius of the granules can be calculated, employing the value of ρ already obtained. In a carefully prepared uniform suspension the three methods gave the following figures for the radii of the granules: 0.371, 0.3667, and 0.3675×10^{-4} cm, respectively.

To determine the ratio of the number of particles at two different levels, an ingenious method was employed. An opaque disc with a small hole pierced in it was placed in the focal plane of the microscope, thus limiting the field of vision so that not more than five or six granules were present at any moment. In this way the exact number could be ascertained by instantaneous observation without deliberate counting. The total count for 200 readings at each of two levels at a known distance h apart, represented by the change in the focus of the microscope, is equivalent to counting the number of granules in an area 200 times as great as that actually under observation. Owing to the movement of the particles it would be impossible to make a direct count of the number in such an area. The ratio of the two totals gives the ratio of the number of granules per unit volume at two levels, which can be inserted in (27).

All the data for the determination of N in (27) were thus available, and from a series of experiments made with mastic and gamboge suspensions in water, at temperatures varying from -9° to 60° C., and in a glycerin-water mixture, so that the apparent density of the granules varied five-fold and the viscosity of the medium over a hundred-fold, the values were found to be almost constant, lying between 6.5×10^{23} and 7.2×10^{23} . In spite of the complete difference in method, this is of the same order as the results obtained by procedures based directly on the

kinetic theory of gases, described below, and from observations of an entirely different type (pp. 13, 122, 129).

Limitation of the Distribution Equation.—E. F. Burton (1921) and A. W. Porter (1922) have shown that the distribution equation (27) is only applicable for the uppermost layers of suspensions where the "concentration" is relatively small. At lower levels the number of granules per unit area approaches a constant value; this is attributed to the fact that in these circumstances the simple kinetic theory cannot be applied. Forces of attraction and repulsion between the granules are important and their actual volume is appreciable. The measurements of Perrin were made in the layers for which the vertical distribution equation should hold, at least approximately.

The Displacement of Particles in Suspensions.—A more direct treatment of the Brownian movement and the assumption of the applicability of the kinetic theory to microscopic granules, provide another method for the determination of the Avogadro number. Although a particle in a suspension is undergoing haphazard movement in all directions, it is possible to consider the displacement parallel to a particular axis in a given time interval τ . The displacement will vary from one granule to another in this time and will differ for the same granule for successive equal intervals of time. As there should be no preference for the movement of the particles in any one direction, it is convenient to assume an average displacement x for all the granules in the time interval τ . It may be supposed that half the molecules move to the right, parallel to the axis under consideration, and the other half move to the left. Consider a horizontal cylinder of 1 sq. cm. cross section containing a suspension the concentration of which is not uniform (Fig. 44). If AA is a reference plane and LL and MM

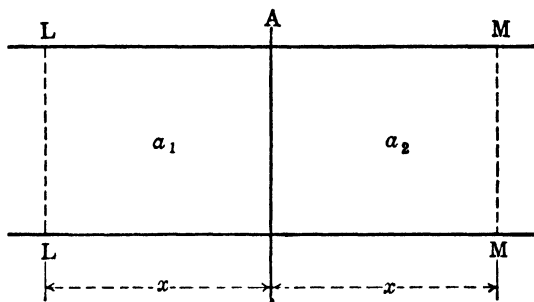


Fig. 44. Displacement of particles in a suspension

represent two planes at a distance x from AA , where x , as already indicated, is the mean displacement of a particle along the axis of the cylinder in the time τ , then on the average half the granules contained in the volume $LLAA$ will have crossed the plane AA from left to right, whereas half the granules in $MMAA$ will have crossed AA from right to left in the given time interval. Any granules outside the range bounded by the planes LL and MM will be at a greater distance from the plane AA than the mean displacement, and so will not be able to reach it. If a_1 and a_2 are the average concentrations of the granules per unit volume in the sections $LLAA$ and $MMAA$, respectively, a_1 being greater than a_2 , the net transfer of granules from left to right is equal to $\frac{1}{2}xa_1 - \frac{1}{2}xa_2$, or

$\frac{1}{2}x(a_1 - a_2)$. It should be noted that since the cylinder has a cross-sectional area of unity, the volumes under consideration are each x cc. If the concentration changes uniformly from LL to MM , the concentration gradient, that is the fall in concentration between two planes at a unit distance apart, across the plane AA may be represented as $(a_1 - a_2)/x$; the situation is equivalent to concentrations of a_1 and a_2 at points midway between L and A , and A and M , respectively, which would thus be at a distance x apart. It is therefore possible to write for the concentration gradient

$$- da/dx = (a_1 - a_2)/x \quad (30)$$

or

$$a_1 - a_2 = - x da/dx, \quad (31)$$

the negative sign being used because the gradient is actually negative, that is the concentration decreases from left to right. Inserting the value for $a_1 - a_2$ in the expression $\frac{1}{2}x(a_1 - a_2)$ obtained above for the amount of material carried across the plane AA in time τ , this quantity becomes $-\frac{1}{2}x^2(da/dx)$; it follows, therefore, that

$$\text{Amount carried across plane of area 1 sq. cm. in unit time} = -\frac{x^2}{2\tau} \cdot \frac{da}{dx}. \quad (32)$$

The diffusion coefficient D is defined as the weight of material passing in unit time across a plane of 1 sq. cm. area when the concentration gradient, $- da/dx$, is unity; putting this value into (32) it is seen that

$$D = \frac{x^2}{2\tau}. \quad (33)$$

It is now necessary to find a relationship between D and other quantities which are directly measurable (A. Einstein, 1905, 1908). Again, consider a horizontal cylinder having a cross-sectional area of 1 sq. cm., containing a suspension, and suppose there are n granules per unit volume at a given plane, and $n - dn$ granules at a distance dx to the right (Fig. 45). If the treatment

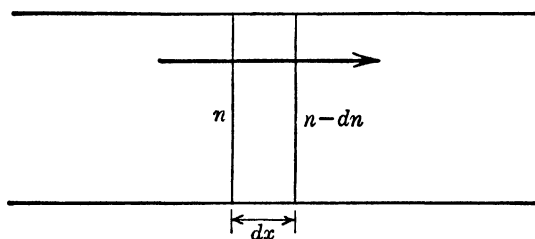


FIG. 45. Force acting on suspended particles

of the kinetic theory of gases may be applied to the suspended particles, then the pressure due to their impacts at the first plane is given by

$$p = \frac{1}{3}mnc^2, \quad (34)$$

and at a distance dx to the right it is

$$p - dp = \frac{1}{3}m(n - dn)c^2, \quad (35)$$

and so the difference of pressure is

$$dp = \frac{1}{3}mc^2dn, \quad (36)$$

where m is the mass of a granule and \bar{c}^2 is the mean square speed of the Brownian movement. The pressure gradient is dp/dx , and since the cross-sectional area is 1 sq. cm., this also represents the force gradient which tends to drive granules from the region of higher to that of lower concentration, that is from left to right in the cylinder. This force may be regarded as acting on the n particles present in unit volume and so the driving force f for a single granule is equal to $(dp/dx)/n$; hence,

$$f = \frac{m\bar{c}^2}{3n} \cdot \frac{dn}{dx}. \quad (37)$$

In general, when a force f acts on a particle it will move with a velocity u given by the relationship

$$u = f/F, \quad (38)$$

where F is the frictional resistance of the medium on the moving granule; therefore in the case under consideration

$$u = \frac{m\bar{c}^2}{3nF} \cdot \frac{dn}{dx}. \quad (39)$$

This gives the average speed with which each granule diffuses from left to right, and the total quantity of material diffusing in unit time is obtained on multiplying the concentration of the granules by their average speed. Since there are n granules per unit volume, the concentration expressed in the equivalent of moles per unit volume is n/N , where N is the Avogadro number; hence the quantity diffusing in unit time is un/N , that is

$$\frac{un}{N} = \frac{m\bar{c}^2}{3FN} \cdot \frac{dn}{dx} \quad (40)$$

$$= \frac{m\bar{c}^2}{3F} \cdot \frac{da}{dx}, \quad (41)$$

since dn/N is equal to da , the difference in concentration between the planes dx apart. This quantity diffuses across 1 sq. cm., the cross-sectional area of the cylinder, in unit time, and so if the concentration gradient da/dx is just equal to unity the diffusion coefficient D is obtained; thus,

$$D = m\bar{c}^2/3F. \quad (42)$$

According to Perrin's assumption, already considered, the kinetic energy of Brownian movement of a suspended granule, $\frac{1}{2}m\bar{c}^2$, is equal to that of a molecule as determined by the kinetic theory, that is to E_K/N ; and since E_K is equal to $\frac{3}{2}RT$ [equation (16)], it follows on substitution in (42) that

$$D = RT/NF. \quad (43)$$

Finally, according to Stokes's law which, as previously mentioned, Perrin showed to be applicable to the particles he studied, the frictional resistance F is given by

$$F = 6\pi\eta r, \quad (44)$$

where η is the viscosity of the medium and r the mean radius of the particles, and so it follows that

$$D = RT/6\pi\eta rN. \quad (45)$$

From the two equations (33) and (45) for the diffusion coefficient there is obtained

$$\frac{x^2}{2\tau} = \frac{RT}{6\pi\eta rN} \quad (46)$$

or

$$N = \frac{RT\tau}{3\pi\eta r x^2}, \quad (47)$$

thus giving an equation from which the Avogadro number may be evaluated.

The only quantity in (47) remaining to be determined is the mean displacement x of a granule along a given axis in the time τ . Perrin found, in agreement with expectation, that the actual displacements followed a Maxwell distribution (see p. 264) and in order to obtain an accurate average it was necessary to take a large number of successive measurements. The method employed was as follows: a single granule was focussed under the microscope and its position at intervals of 30 sec. (τ) was plotted. The distances between successive positions as plotted on paper (Fig. 46A) give the displacement l

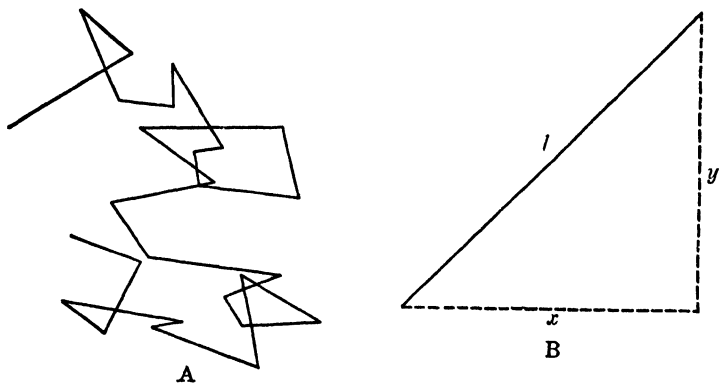


FIG. 46. Displacements of a single granule

projected on to a horizontal plane, and this may be split up into two components x and y along axes at right angles (Fig. 46B) so that $x^2 + y^2 = l^2$. For a large number of displacements it follows that

$$\Sigma x^2 + \Sigma y^2 = \Sigma l^2, \quad (48)$$

and as the movement is irregular, Σx^2 should be equal to Σy^2 , and hence

$$\Sigma x^2 = \frac{1}{2} \Sigma l^2. \quad (49)$$

The quantity x^2 required in (47) is the mean of the terms represented by Σx^2 , and so the former can be put equal to half the mean-square of the distances l between successive positions on a horizontal plane at 30 sec. intervals. The mean should preferably be taken for a number of particles.

By this method, using suspensions of both gamboge and mastic, in which the masses of the granules varied in the ratio of about 15,000 to 1, and with viscosity variations of 125 to 1 in different cases, the value of N

given by (47) was found by Perrin to be between 5.5×10^{23} and 8.0×10^{23} , the most reliable value being about 6.6×10^{23} . Other workers have used the same method with colloidal particles of gold (T. Svedberg, 1911) and of mercury (I. Nordlund, 1914), and with suspensions of bacteria (J. H. Shaxby, 1923) which also show Brownian movement. In some cases (45) has been used to determine N , a direct measurement of D being made by observing the total quantity of material diffusing across a boundary under given conditions. Colloidal solutions of gold (T. Svedberg, 1911; A. Westgren, 1914), and suspensions of gamboge (L. Brillouin, 1912) and of bacteria (J. H. Shaxby, 1925) have been investigated. Observations have also been made of average displacements due to the Brownian movement of particles suspended in air, the Einstein formula deduced above being applied to calculate the Avogadro number; particles of oil and of selenium were used. In all these experiments the value of N was found to be about 6×10^{23} .

Relatively large suspended particles show a rotational, in addition to a translational, Brownian movement, and the angular displacements can be related to the Avogadro number by an equation analogous to that for translational motion. The experimental difficulties in making the necessary measurements are considerable; nevertheless J. Perrin (1909) succeeded in studying a suspension of gamboge containing spherical particles of 13×10^{-4} cm. diameter. The results obtained for the Avogadro number by the various methods described, depending on the study of suspended particles, are summarized in Table 31; the range given in each case covers the best values of different authors.

TABLE 31. THE AVOGADRO NUMBER FROM BROWNIAN MOVEMENT

Experimental Method	Avogadro Number
Vertical Distribution of Granules	$6.0-7.0 \times 10^{23}$
Translational Displacement in Liquids	5.9-7.2
Translational Displacement in Gases	5.7-6.0
Diffusion of Granules	4.4-6.5
Rotational Displacement in Liquid	6.5

Considering the experimental difficulties and the magnitude of the figures involved, these results are remarkably constant, but the striking fact is that they are in good agreement with estimates of the Avogadro number made by entirely different methods. The proximity of the value to that obtained from the mean free path of a gas and the density of a liquid at low temperatures, to be considered shortly, may be taken as no more than proof that Perrin's analogy between a gas and a suspension is justified, but the fact that the same result may be arrived at by entirely different methods not involving the kinetic theory, must be regarded as providing evidence for that theory and for the real existence of molecules.²

The Avogadro Number.—The most accurate value of the Avogadro number is probably that based on the direct measurement of the wave

length of X-rays; the result obtained in this way (p. 13), viz., 6.023×10^{23} , will be adopted here. A knowledge of the Avogadro number is important in many aspects of physical chemistry. The weight of a single atom or molecule of any substance can be determined, for example, by dividing the corresponding conventional atomic or molecular weight by 6.023×10^{23} . A molecule of hydrogen thus weighs 3.346×10^{-24} g., and the weight of a single hydrogen atom is 1.673×10^{-24} g.

Reference has been made earlier to the gas constant per single molecule, generally known as the **Boltzmann constant** k , which is equal to R/N , i.e., the molar gas constant divided by the Avogadro number. The value is usually expressed in erg deg.⁻¹ units, and since R is 8.314×10^7 ergs deg.⁻¹ mole⁻¹, it follows that

$$k = \frac{R}{N} = \frac{8.314 \times 10^7}{6.023 \times 10^{23}} = 1.380 \times 10^{-16} \text{ erg deg.}^{-1}$$

Maxwell's Law of the Distribution of Velocities.—As a result of impacts, the molecules of a gas cannot all have the same velocity (p. 248), and it is of interest to know how the velocities are distributed between the molecules when a steady state is attained. The term "steady state" is not meant to imply that all the molecules retain their velocities unchanged, but rather that the *distribution* of velocities is unchanged, although as a result of impacts energy and velocity exchanges between the molecules are proceeding continually. The treatment outlined below is based on that first given by J. Clerk Maxwell (1860).

Consider a molecule with velocity c , the components of which along three axes at right angles are u , v and w , so that, as already seen (p. 249),

$$c^2 = u^2 + v^2 + w^2. \quad (50)$$

If the assumption is made that the values of u , v and w for any molecule are completely independent of each other,* then the probability that the u component lies between u and $u + du$ is a function of u alone, i.e., $f(u)du$, and similarly for the other velocity components. The probability W that the three components of the molecular velocity shall simultaneously lie between u and $u + du$, between v and $v + dv$, and between w and $w + dw$, respectively, is then given by

$$W = f(u)f(v)f(w)dudvdw. \quad (51)$$

Suppose that while the total velocity c is maintained constant, the individual components are varied by the infinitesimal amounts δu , δv and δw , respectively; this is equivalent to making a small change in the direction of motion without affecting the speed, i.e., δc is zero. The condition that δc , and hence δc^2 , is zero is obtained from (50); thus,

$$\begin{aligned} \delta c^2 &= \delta(u^2 + v^2 + w^2) = 0, \\ \therefore u\delta u + v\delta v + w\delta w &= 0. \end{aligned} \quad (52)$$

Since the motion of the molecules is completely random, no direction being preferred over any other, it is evident that the functions $f(u)$, $f(v)$ and $f(w)$,

* This assumption appears to be open to criticism; nevertheless, the same final result may be obtained by methods which are free from this objection.

which determine the numbers of molecules with components in given ranges, will remain constant, provided c is constant; hence the variation in the product $f(u)f(v)f(w)$ will be zero, viz.,

$$\delta[f(u)f(v)f(w)] = 0, \\ \therefore f'(u)f(v)f(w)\delta u + f(u)f'(v)f(w)\delta v + f(u)f(v)f'(w)\delta w = 0, \quad (53)$$

where $f'(u)$ represents the derivative of $f(u)$ with respect to u , i.e., $df(u)/du$, and so on. If (53) is divided by $f(u)f(v)f(w)$, it follows that

$$\frac{f'(u)}{f(u)}\delta u + \frac{f'(v)}{f(v)}\delta v + \frac{f'(w)}{f(w)}\delta w = 0. \quad (54)$$

Equation (52) is now multiplied by an arbitrary factor x , and the resulting quantity is then added to (54), giving

$$\left[\frac{f'(u)}{f(u)} + xu \right] \delta u + \left[\frac{f'(v)}{f(v)} + xv \right] \delta v + \left[\frac{f'(w)}{f(w)} + xw \right] \delta w = 0. \quad (55)$$

Since the variations δu , δv and δw are completely arbitrary, for a given velocity c , it follows that each of the quantities in the brackets must be zero; that is,

$$\frac{f'(u)}{f(u)} + xu = 0 \quad \text{or} \quad \frac{f'(u)}{f(u)} = -xu. \quad (56)$$

In view of the fact that $f'(u)$ is the first derivative of $f(u)$, it is apparent that $f'(u)/f(u)$ is equivalent to $d \ln f(u)$, and consequently integration of (56) gives

$$\ln f(u) = -\frac{xu^2}{2} + \ln A, \quad (57)$$

where $\ln A$ is the integration constant. This equation may be written in the form

$$f(u) = Ae^{-xu^2/2}, \quad (58)$$

or alternatively, replacing $x/2$ for convenience by a quantity $1/\alpha^2$, the expression becomes

$$f(u) = Ae^{-u^2/\alpha^2}. \quad (59)$$

Treating the other terms in (55) in an analogous manner, it follows that

$$f(v) = Ae^{-v^2/\alpha^2} \quad \text{and} \quad f(w) = Ae^{-w^2/\alpha^2}, \quad (59a)$$

and hence (51) becomes

$$W = A^3 e^{-(u^2+v^2+w^2)/\alpha^2} du dv dw \\ = A^3 e^{-c^2/\alpha^2} du dv dw. \quad (60)$$

In order to make (60) of direct use it is necessary to evaluate A and α ; this may be done in the ensuing manner. Since the velocity components can have all values between $+\infty$ and $-\infty$, it follows that the probability of a value of u between these limits must be unity. In other words, the integral of $f(u)du$ between $+\infty$ and $-\infty$ must be equal to unity, i.e.,

$$\int_{-\infty}^{+\infty} Ae^{-u^2/\alpha^2} du = 1, \quad (61)$$

and hence, by carrying out the actual integration,

$$A\pi^{\frac{1}{2}}\alpha = 1 \quad (62)$$

or

$$A = 1/\pi^{\frac{1}{2}}\alpha. \quad (63)$$

Substitution in (60) then gives

$$W = \frac{1}{\pi^{\frac{1}{2}}\alpha^3} e^{-c^2/\alpha^2} du dv dw, \quad (64)$$

or transferring rectangular to polar coordinates,

$$W = \frac{1}{\pi^{\frac{1}{2}}\alpha^3} e^{-c^2/\alpha^2} c^2 dc \sin \theta d\theta d\phi. \quad (65)$$

This equation gives the probability that the velocity of a molecule will lie between c and $c + dc$, the *direction* of c lying between the limits θ and $\theta + d\theta$, and ϕ and $\phi + d\phi$. In order to obtain the fraction of the total number of molecules, that is dn_c/n , having velocities between c and $c + dc$, irrespective of direction, it is necessary to integrate (65) with respect to θ between the limits 0 and π , and with respect to ϕ between the limits 0 and 2π , thus including all possible directions. It follows then, that

$$\frac{dn_c}{n} = \frac{4}{\pi^{\frac{1}{2}}\alpha^3} e^{-c^2/\alpha^2} c^2 dc. \quad (66)$$

The significance of α may be seen by expressing the mean square velocity \bar{c}^2 in the form

$$\bar{c}^2 = \frac{\int_0^\infty c^2 dn_c}{n}, \quad (67)$$

and then replacing dn_c/n by the right-hand side of (66); that is,

$$\bar{c}^2 = \frac{4}{\pi^{\frac{1}{2}}\alpha^3} \int_0^\infty e^{-c^2/\alpha^2} c^4 dc. \quad (68)$$

The value of the integral is $3\pi^{\frac{1}{2}}\alpha^5/8$, so that \bar{c}^2 is $3\alpha^2/2$, and

$$\alpha^2 = 2\bar{c}^2/3. \quad (69)$$

It has been seen (p. 251) that $PV = Nm\bar{c}^2/3$; for a mole of gas N is the Avogadro number, and PV is then equal to RT , and so

$$\frac{1}{2}Nm\bar{c}^2 = RT, \quad (70)$$

which combined with (69) gives

$$\alpha^2 = \frac{2RT}{Nm} = \frac{2kT}{m}, \quad (71)$$

where k , equal to R/N , is the Boltzmann constant, i.e., the gas constant per single molecule. Substituting the value of α^2 obtained from (71) in (66) it is

seen that

$$\frac{dn_c}{n} = 4\pi \left(\frac{m}{2\pi kT} \right)^{\frac{3}{2}} e^{-mc^2/2kT} c^2 dc; \quad (72)$$

$$\therefore \frac{1}{n} \cdot \frac{dn_c}{dc} = 4\pi \left(\frac{m}{2\pi kT} \right)^{\frac{3}{2}} e^{-mc^2/2kT} c^2. \quad (73)$$

If numerator and denominator of the fraction m/k are multiplied by N , the result mN/kN is equal to M/R , where M is the molecular weight of the gas; hence (73) may be rewritten as

$$\frac{1}{n} \cdot \frac{dn_c}{dc} = 4\pi \left(\frac{M}{2\pi RT} \right)^{\frac{3}{2}} e^{-Mc^2/2RT} c^2 \quad (74)$$

If E represents the kinetic energy of 1 mole of molecules having the velocity c , i.e., $E = Mc^2/2$, it can be readily shown that

$$M^{\frac{1}{2}} c^2 dc = (2E)^{\frac{1}{2}} dE,$$

and substitution in (74) leads to the result

$$\frac{1}{n} \cdot \frac{dn_c}{dE} = \frac{2\pi}{(\pi RT)^{\frac{3}{2}}} e^{-E/RT} E^{\frac{1}{2}}. \quad (75)$$

The equations (72) to (75) are different forms of Maxwell's law of the distribution of molecular velocities. The significance of this law may be seen by a graphical representation of these equations; in Fig. 47 the quantity $\frac{1}{n} \cdot \frac{dn_c}{dc}$, evaluated from the right-hand side of (73) or (74), is plotted as ordinate against various values of the velocity c as abscissae, at constant temperature.* The area between two ordinates separated by dc is clearly equal to

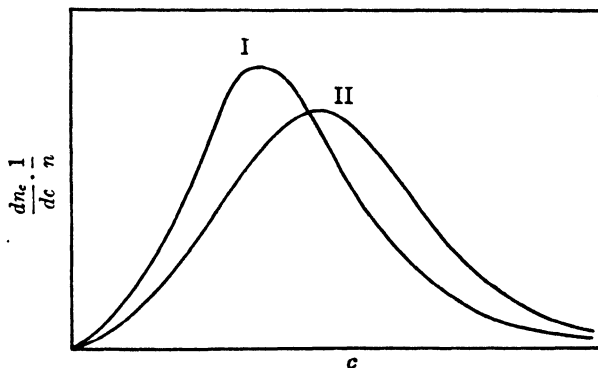


FIG. 47. Maxwell's distribution law

dn_c/n , that is to the fraction of the total number of molecules possessing velocities between c and $c + dc$; hence the area under the whole curve is equal to the total number of molecules, and the height of the ordinate corresponding to any velocity is virtually a measure of the fraction of the molecules which have

* A curve of a similar type would be obtained if c^2 or E were plotted as abscissae.

that velocity. The distribution law shows, therefore, that the majority of molecules have speeds lying within a relatively limited range, but a certain proportion of the molecules have very low and very high velocities. By altering the temperature in (73), (74) or (75) the general shape of the distribution curve is unchanged, but increase of temperature results in a flattening of the maximum, which now occurs at a higher velocity. In Fig. 47, for example, curve I represents the distribution of velocities at a lower and II a higher temperature. It is seen, therefore, that as the temperature is raised there is a wider distribution of velocities, and the number of molecules possessing high speeds is increased; this effect is mainly due to the presence of the exponential term which increases rapidly with increasing temperature.

The Most Probable Velocity.—If the right-hand side of (73) is differentiated with respect to c and the result equated to zero, the conditions for the maxima and minima in the distribution curve (Fig. 47) are obtained. The relationship found is

$$\left(1 - \frac{mc^2}{2kT}\right) ce^{-mc^2/2kT} = 0. \quad (76)$$

Two possible solutions of this equation are for c to be 0 or ∞ ; these clearly correspond to minima, for there are no molecules having velocities of zero or infinity. A third solution is for the quantity in the parentheses to be equal to zero; that is,

$$1 - \frac{mc^2}{2kT} = 0; \\ \therefore c = \sqrt{2kT/m}. \quad (77)$$

It was shown above that k/m is equal to R/M , so that

$$c = \sqrt{2RT/M}. \quad (78)$$

This gives the condition for the maximum of the distribution curve, and the corresponding value of c is, therefore, the **most probable velocity**; more molecules have this velocity than any other single value. By comparison with (18) it is clear that the most probable velocity is $\sqrt{2/3}$ times the square root of the mean square velocity.

The Mean Velocity.—The mean velocity may be expressed by an equation similar to (67),

$$\bar{c} = \frac{\int_0^\infty c dn_c}{n}, \quad (79)$$

and replacing dn_c by the value obtained from (72) and integrating, it is found that

$$\bar{c} = \sqrt{8kT/\pi m} = \sqrt{8RT/\pi M}. \quad (80)$$

Comparison with (18) shows that

$$\bar{c} = \sqrt{8/3\pi} \sqrt{\bar{c}^2} = 0.921 \sqrt{\bar{c}^2}, \quad (81)$$

thus bringing out clearly the difference between the mean velocity and the square root of the mean square velocity.

Experimental Verification of the Distribution Law.—In the method used by Stern (p. 254) to determine the velocity of silver atoms, it was observed that the values were spread over a range, in general agreement with the requirements of the distribution law. By improvements of technique it has been possible to determine, approximately, the shape of the velocity distribution curve. J. L. Costa, H. D. Smyth and K. T. Compton (1927) mounted two toothed discs, at a definite distance apart, on a common axle; a beam of molecules, e.g., hydrogen, nitrogen or carbon tetrachloride, was arranged so that when the discs were stationary the beam passed through the gaps between the teeth on to a radiometer vane, the deflection of the latter giving a measure of the number of molecules striking it. The discs were then set in rotation. A molecule passing through a gap in the first will pass through a gap in the second only if the time required to travel the distance between the discs is equal to an integral multiple of the time necessary for the discs to rotate from one gap to the next. From a knowledge of the speed of the discs, therefore, the velocity of the molecules passing through can be determined, and the number possessing this velocity is given by the deflection of the radiometer vane. By making measurements with the discs rotating at various rates, it is possible to obtain an approximate velocity distribution curve. A method involving a somewhat similar principle has been employed by J. A. Eldridge (1927).

The procedure adopted by I. F. Zartman (1931) and C. C. Ko (1934) is analogous to that of Stern (p. 254). A beam of metal atoms or molecules, e.g., bismuth, is sent out from a heated filament, and passing through fine slits falls on a rotating drum with a similar slit. Once in every rotation the latter slit comes opposite the beam of molecules and these enter the drum; owing to the differences in their velocities they do not all strike the opposite side of the drum at the same point, but are spread out into a band. Measurement of the intensity of the band by means of a photometer gives an indication of the distribution of velocities.

In the method of W. Meissner and H. Scheffers (1933) a beam of lithium or potassium atoms, after passing through a slit, traverses an inhomogeneous magnetic field, similar to that used in the Stern-Gerlach experiment (p. 64). Each deflected beam is spread out and the intensity at various points depends on the number of molecules reaching that point. The positive ion current resulting from the alkali metal atoms striking a fine, incandescent platinum wire parallel to the slit is observed, and this gives a measure of the number of atoms in that position, and consequently the number having a particular speed. The wire is moved by a micrometer, and the number of atoms determined in a series of positions so as to obtain the velocity distribution.

All the methods described show that the Maxwell distribution law is obeyed, under the conditions of measurement, within the limits of experimental error.

Simplified Distribution Law.—In connection with a theory of reaction rates (p. 1090) it is required to know the proportion of molecules having energy in excess of a specified value; this may be obtained from the distribution equation by integration, but the resulting expression is complicated. The simplified treatment given below, however, gives a

result which may be regarded as approximately true. The assumption made is that the molecular movements are restricted to a plane, so that there are only two, instead of three, rectangular components of the velocities. The expression for W , the probability of a molecule of velocity c having simultaneously components lying between u and $u + du$, and between v and $v + dv$, is [compare equation (64)] then

$$W = \frac{1}{\pi\alpha^2} e^{-c^2/\alpha^2} du dv, \quad (82)$$

or, in polar coordinates,

$$W = \frac{1}{\pi\alpha^2} e^{-c^2/\alpha^2} c dc d\theta. \quad (83)$$

Integrating for values of θ lying between 0 and 2π , that is over all directions in the plane, it is found that

$$\frac{dn_c}{n} = \frac{2}{\alpha^2} e^{-c^2/\alpha^2} c dc, \quad (84)$$

and since α^2 has the same value as previously, i.e., $2kT/m$ or $2RT/M$, it follows that

$$\frac{dn_E}{n} = \frac{1}{RT} e^{-E/RT} dE, \quad (85)$$

where E is the kinetic energy per mole for molecules having the same speed c . The fraction of the total number of molecules having energy in excess of a given value E is obtained by integration; thus,

$$\frac{n_E}{n} = \frac{1}{RT} \int_E^\infty e^{-E/RT} dE = e^{-E/RT}, \quad (86)$$

where n_E is the number of molecules possessing translational energy, in two components, in excess of E per mole and n is the total number of molecules.

The Maxwell-Boltzmann Distribution Law.—Maxwell's law of the distribution of molecular velocities is an aspect of a more general law which is particularly associated with the names of J. C. Maxwell (1860) and L. Boltzmann (1868). The treatment, based on the theory of probability, leads to an expression for the number of molecules in a gas at equilibrium possessing a certain specified energy. An exact discussion requires the use of the principles of statistical mechanics, and these will be tacitly assumed here.

Consider a system made up of n similar, but *distinguishable*, molecules which do not attract each other, e.g., as in an ideal gas. If n_1 molecules each possess energy ϵ_1 , n_2 molecules have energy ϵ_2 , and so on, the total number n is the sum of $n_1, n_2, \dots, n_i, \dots$, while the total energy E is the sum of the terms $n_1\epsilon_1, n_2\epsilon_2, \dots, n_i\epsilon_i, \dots$; thus,

$$\sum n_i = n \quad \text{and} \quad \sum n_i\epsilon_i = E. \quad (87)$$

If the total number of molecules and the total energy of the system are constant, the variations δn and δE must be set equal to zero, so that, using (87),

$$\delta n = \sum_i \delta n_i = 0 \quad (88)$$

and

$$\delta E = \sum_i \epsilon_i \delta n_i = 0. \quad (89)$$

The **probability** W of the given system may be taken as proportional to the number of different ways in which the molecules can be distributed, under the specified conditions. This can be calculated by finding the number of ways in which n distinguishable objects can be arranged among a set of boxes, representing different energy values, there being always n_1 in the first, n_2 in the second, and so on. The required probability is then

$$W = \frac{n!}{n_1! n_2! \cdots n_i! \cdots} \times \text{constant}, \quad (90)$$

and hence

$$\ln W = \ln n! - \sum_i \ln n_i! + \text{constant}. \quad (91)$$

This expression may be simplified by using Stirling's formula for the factorials of large numbers, viz., $\ln n! \approx n \ln n - n$, so that (91) becomes

$$\ln W = n \ln n - \sum_i n_i \ln n_i + \text{constant}, \quad (92)$$

remembering that $\sum n_i$ is equal to n by (87).

The actual (equilibrium) state of the system may be taken as that having the maximum probability, that is, when W is a maximum. Under these conditions the variation δW (or $\delta \ln W$) is zero, so that from (92)

$$\begin{aligned} \delta \ln W &= - \sum_i \delta(n_i \ln n_i) = 0; \\ \therefore \sum_i (\ln n_i + 1) \delta n_i &= 0. \end{aligned} \quad (93)$$

The equations (88), (89) and (93) represent the conditions which must be satisfied simultaneously for the system of maximum probability. If (88) and (89) are multiplied by the undetermined constants α and β , respectively, and the results added to (93), it is found that

$$\sum_i (\ln n_i + 1 + \alpha + \beta \epsilon_i) \delta n_i = 0. \quad (94)$$

The integer 1 may be combined with α to give a new constant a , so that (94) may be written as

$$\sum_i (\ln n_i + a + \beta \epsilon_i) \delta n_i = 0.$$

Since the variations $\delta n_1, \delta n_2, \dots, \delta n_i, \dots$, are independent of each other, provided the conditions of (88) and (89) are satisfied, it follows that every

term in this summation must be zero, so that

$$\begin{aligned}\ln n_i + a + \beta\epsilon_i &= 0; \\ \therefore \ln n_i &= -(a + \beta\epsilon_i) \\ n_i &= Ce^{-\beta\epsilon_i},\end{aligned}\tag{95}$$

where C , equal to e^{-a} , is also a constant. The value of β is found by utilizing (95) to derive expressions for various thermodynamic quantities, e.g., energy or pressure, of an ideal gas, and comparing them with the results obtained in other ways. In this manner β can be shown to be equal to $1/kT$, where k is the Boltzmann constant and T is the absolute temperature, so that (95) becomes

$$n_i = Ce^{-\epsilon_i/kT}.\tag{96}$$

This result may be taken as the expression of the **Maxwell-Boltzmann (or classical) distribution law**; it gives the number of molecules n_i possessing energy ϵ_i at any temperature T . In deriving (96) no restriction was made as to the nature of the energy, and so it may be regarded as applying to the distribution of the total energy or of any individual form of energy, e.g., translational, vibrational, rotational, etc.

The classical distribution law equation (96) has numerous applications in physical chemistry, and some of these will be given here. If it is required to determine the fraction n_i/n of the total number of molecules which have the particular energy ϵ_i , use may be made of (87) in conjunction with (96); thus,

$$n = \sum_i n_i = C \sum_i e^{-\epsilon_i/kT},$$

and hence, from (96),

$$\frac{n_i}{n} = \frac{e^{-\epsilon_i/kT}}{\sum_i e^{-\epsilon_i/kT}}.\tag{97}$$

Suppose n_1 and n_2 are the numbers of molecules possessing energies ϵ_1 and ϵ_2 , respectively; then from (97)

$$\frac{n_2}{n_1} = \frac{e^{-\epsilon_2/kT}}{e^{-\epsilon_1/kT}} = e^{-\Delta\epsilon/kT}\tag{98}$$

where $\Delta\epsilon$, equal to $\epsilon_2 - \epsilon_1$, is the energy difference per molecule. If ΔE , equal to $N\Delta\epsilon$, is the energy difference per mole, N being the Avogadro number, then (98) may be written as

$$\frac{n_2}{n_1} = e^{-\Delta E/RT},\tag{98a}$$

since Nk is equal to R .

A particular application of (98) or (98a) is in connection with the distribution of molecules in a nonuniform force field, in which the potential energies vary with the position. If the energy of a molecule is separable into two independent parts, viz., the kinetic energy, determined by the speed, i.e., temperature, and the potential energy, then for a system in equilibrium at constant temperature in a nonuniform force field, the energy differences of the molecules will be those of the potential energies only. If n_A and n_B represent the average numbers of molecules per unit volume at points A and B, where the potential

energies differ by u per molecule or U per mole, then by (98) or (98a),

$$\frac{n_A}{n_B} = e^{-u/kT} = e^{-U/RT}, \quad (99)$$

the energy being greater at A.*

For certain purposes (pp. 32, 332) it is convenient to express the energy in any level with reference to that in the lowest, i.e., zero-point, level. If in (98) the number of molecules n_1 is set equal to n_0 , the number in the lowest level, then ϵ_1 may be taken as zero, and hence n_2/n_0 becomes equal to $e^{-\epsilon_2/kT}$; in general, therefore,

$$n_i = n_0 e^{-\epsilon_i/kT}, \quad (100)$$

where n_i is the number of molecules in the state for which the energy per molecule is ϵ_i in excess of the zero-point value. Actually (100) is generally employed in the form

$$n_i = p_i n_0 e^{-\epsilon_i/kT} \quad \text{or} \quad N_i = p_i N_0 e^{-\epsilon_i/kT}, \quad (101)$$

where p_i , the **statistical weight factor** (p. 95), is introduced to allow for the possibility that there may be a number of quantum levels of almost identical energy. As will be seen shortly, quantum mechanics requires the inclusion of this factor in all forms of the Maxwell-Boltzmann distribution law; strictly speaking, therefore, (96) should be written $n_i = p_i C e^{-\epsilon_i/kT}$.

It is possible to derive Maxwell's equation for the distribution of velocities from the general expression (96). Several methods have been used, but as they are relatively complicated they will not be given here. It will be sufficient to point out that the exponential factor, which is the significant one, appears in (72), etc., as well as in (96).

Quantum Statistics.—From the point of view of modern quantum mechanics, it is not possible for similar particles to be distinguishable; the fundamental assumption of the Maxwell-Boltzmann (classical) statistics (p. 270) is consequently unsound. In the quantum statistics, therefore, it is postulated that the particles are not distinguishable. In the form developed by S. N. Bose (1924) and A. Einstein (1925), which is applicable to photons, i.e., radiation (p. 33), and to atoms or molecules containing an even number of protons, neutrons and electrons, it is supposed that the eigenfunction of the system is *symmetric* (cf. p. 79). In other words, according to the **Bose-Einstein statistics**, an interchange of the coordinates of any two particles will produce no change in the sign of the eigenfunction. The number of individual wave functions associated with each energy state ϵ_i is equal to the statistical weight p_i , and in the Bose-Einstein statistics there is no restriction as to the number n_i of indistinguishable particles that may be associated with each wave function. By utilizing a procedure analogous to that described above, the distribution law, for the system of maximum probability, is given by

$$n_i = \frac{p_i}{C^{-1} e^{\epsilon_i/kT} - 1}, \quad (102)$$

n_i being the number of particles possessing energy ϵ_i ; C is a constant, similar to that in (96).

* The hypsometric formula (25) is a form of (99), since E_K is equal to $\frac{1}{2}RT$.

In the statistics of E. Fermi (1926) and P. A. M. Dirac (1927), applicable to electrons, in particular, and to atoms and molecules containing an odd number of fundamental particles, the eigenfunction of the system is taken as being *antisymmetric*. In the **Fermi-Dirac statistics**, therefore, an interchange in the coordinates of two particles results in a change in sign of the eigenfunction. In these circumstances, only one of the n_i particles of energy ϵ_i can be associated with each of the p_i wave functions. It is then found that

$$n_i = \frac{p_i}{C^{-1}e^{\epsilon_i/kT} + 1}. \quad (103)$$

Strictly speaking (102) or (103), according to the nature of the particles, should always be used in place of (96), for the distribution of molecules among the possible energy levels. However, except for photons, electrons and liquid helium II (p. 512), the integer 1 in the denominators of (102) and (103) is so small in comparison with $C^{-1}e^{\epsilon_i/kT}$, except at the very lowest temperatures, that it may be neglected. The two equations then reduce to the same form, viz.,

$$n_i = \frac{p_i}{C^{-1}e^{\epsilon_i/kT}} = p_i C e^{-\epsilon_i/kT}, \quad (104)$$

which is identical with the classical distribution equation (96), apart for the introduction of the statistical weight factor. It follows, therefore, in spite of its incomplete basis, that the classical statistics can be applied to systems of atoms and molecules under all reasonable conditions. For electrons, however, the Fermi-Dirac statistics must be used, and as a result important deductions have been made, especially in connection with the properties of the "electron gas" in metals (p. 379).³

The Mean Free Path.—The average distance a molecule moves between two successive collisions is called the **mean free path**: its value may be calculated in the following manner. Suppose all but one of the molecules in a given space are at rest and the moving one continues with

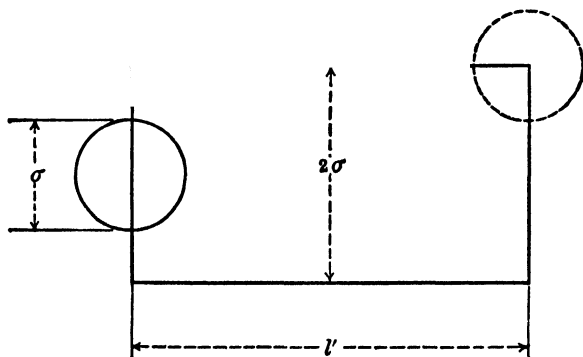


FIG. 48. Molecular collision and mean free path

its normal speed. If l' is the mean free path under these conditions, then on the average the molecule will just collide with another after it has moved through this distance (Fig. 48). When two molecules approach

each other a point is reached at which the repulsion between them is so great as to cause a reversal of direction. Even if a molecule were infinitesimally small, as postulated for an ideal gas, it would have an effective **collision diameter**. This may be regarded as the closest distance of approach between the centers of two molecules taking part in a collision. If σ is the collision diameter of the molecule, then it is evident from the diagram that on the average there can only be one molecule in the cylinder of length l' and diameter 2σ ; that is to say, each molecule inhabits a volume $\pi\sigma^2l'$. For a gas in which there are N molecules in the volume V the volume per molecule is V/N , and so the two values may be equated; thus,

$$\begin{aligned}\pi\sigma^2l' &= V/N; \\ \therefore l' &= V/\pi N\sigma^2.\end{aligned}\tag{105}$$

The mean free path when all the molecules are in motion is different from this value, and the correction may be arrived at by considering the relationship between the actual velocity of a molecule and the velocity *relative* to stationary molecules; it has been shown by J. C. Maxwell that for a mixture of two gases the mean relative velocity $\bar{c}_{\text{rel.}}$ is related to the actual velocities \bar{c}_A and \bar{c}_B by the expression

$$(\bar{c}_{\text{rel.}})^2 = (\bar{c}_A)^2 + (\bar{c}_B)^2.\tag{106}$$

In a single gas \bar{c}_A is equal to \bar{c}_B , and so

$$\bar{c}_{\text{rel.}} = \bar{c}\sqrt{2},$$

the relative velocity being greater than the actual velocity; this means that the velocity of a molecule with respect to stationary molecules is greater than its true velocity. The mean free path l' , as given by (105), will thus be greater than the correct value l in the same proportion, so that l' is equal to $l\sqrt{2}$; that is,

$$\begin{aligned}l &= V/\sqrt{2}\pi N\sigma^2 \\ &= 1/\sqrt{2}\pi n\sigma^2,\end{aligned}\tag{107}$$

n being the number of molecules in unit volume of the gas. The number of molecules in 1 mole is about 6×10^{23} , and 1 mole of any gas occupies approximately 22,400 cc. at 0°C . and 1 atm. pressure, so that n is 2.7×10^{19} . Most gaseous molecules have a diameter of the order of 2×10^{-8} cm., and assuming this value a rough estimate of the mean free path of a gas may be made from (107); thus,

$$l = \frac{1}{\sqrt{2}\pi \times 2.7 \times 10^{19} \times 4 \times 10^{-16}} = 2.1 \times 10^{-5} \text{ cm.}$$

It is apparent, therefore, that under normal conditions the molecules of a gas only travel the very short distance of about 10^{-5} cm. between two successive collisions. The smaller the molecules the larger the mean free

path, but as the diameters of gas molecules do not vary very greatly the mean free path is approximately constant for all gases under the same conditions. As will be seen later, more reliable values can be obtained from viscosity measurements.

Since the mean free path is inversely proportional to the number of molecules in unit volume of the gas [equation (107)], it will also be inversely proportional to the pressure; the mean free path thus increases as the pressure is diminished, the product of the two quantities being constant. It follows, therefore, that a molecule will have a mean free path of 1 cm. at a pressure of approximately 2.1×10^{-5} atm. at 0°C .

Mean Free Path in a Mixture: Molecular Collisions.—It is of interest next to consider a mixture of gases which may be distinguished by the suffixes A and B. Suppose all the molecules except one of type A are fixed; then the average distance which this travels before it collides with a molecule of B is given by

$$l_{A,B} = V/\pi N_B(\frac{1}{2}\sigma_A + \frac{1}{2}\sigma_B)^2, \quad (108)$$

where σ_A and σ_B are the collision diameters of the two kinds of molecules, and N_B is the number of B molecules in the volume V . If all the molecules are in motion it is necessary to divide $l_{A,B}$ by the correction term $\bar{c}_{\text{rel.}}/\bar{c}_A$, which is defined by (106) above; this may be written in the form

$$\bar{c}_{\text{rel.}} = \{(\bar{c}_A)^2 + (\bar{c}_B)^2\}^{\frac{1}{2}},$$

and so

$$\frac{\bar{c}_{\text{rel.}}}{\bar{c}_A} = \left\{ \frac{(\bar{c}_A)^2 + (\bar{c}_B)^2}{(\bar{c}_A)^2} \right\}^{\frac{1}{2}} = \left(\frac{M_A + M_B}{M_B} \right)^{\frac{1}{2}},$$

where M is the molecular weight, since \bar{c} is inversely proportional to $M^{\frac{1}{2}}$ [see equation (80)]. Applying the necessary correction to (108), it follows, therefore, that the average distance traversed by a molecule A in a straight line before it collides with a molecule B is given by

$$l_{A,B} = \frac{4V}{\pi N_B(\sigma_A + \sigma_B)^2} \left(\frac{M_A + M_B}{M_B} \right)^{-\frac{1}{2}}. \quad (109)$$

In order to determine the number of collisions in unit time between one molecule of A and all the molecules of B, the mean velocity (\bar{c}_A) of A is divided by $l_{A,B}$; the total number of collisions in unit time between all the molecules of A and all of B in unit volume ($Z_{A,B}$) will then be obtained by multiplying the result by the number of molecules of A in unit volume, i.e., N_A/V ; hence, using (80),

$$Z_{A,B} = \frac{\bar{c}_A N_A}{l_{A,B} V} = \frac{N_A N_B}{4V^2} (\sigma_A + \sigma_B)^2 \left\{ 8\pi RT \left(\frac{M_A + M_B}{M_A M_B} \right) \right\}^{\frac{1}{2}}. \quad (110)$$

If n_A and n_B are the numbers of molecules per unit volume, $N_A N_B/V^2$ may be replaced by $n_A n_B$, so that

$$Z_{A,B} = \frac{1}{4} n_A n_B (\sigma_A + \sigma_B)^2 \{ 8\pi RT (M_A + M_B) / M_A M_B \}^{\frac{1}{2}} \quad (111)$$

To determine the number of collisions between the molecules of a single gas, (111) may be modified by putting $n_A = n_B$, $\sigma_A = \sigma_B$, and $M_A = M_B$; the result must then be divided by 2, otherwise each collision is counted twice, since all the molecules of A would be treated both as colliding and being collided with. The equation obtained in this way for the number of molecular collisions per second in unit volume of a single gas is

$$Z = 2n^2\sigma^2(\pi RT/M)^{\frac{1}{2}}. \quad (112)$$

The number of molecules entering into collision is $2Z$, since two molecules are involved in each collision.

Some idea of the large number of impacts undergone by the molecules of a gas, supposing the arguments of the kinetic theory to be correct, may be ascertained by inserting approximate values of n and σ in (112). For oxygen, for example, M is 32 whereas σ is about 2×10^{-8} cm., n at 0° c. and 1 atm. pressure is, as seen above, 2.7×10^{19} molecules per cc., and R is 8.3×10^7 ergs deg. $^{-1}$ mole $^{-1}$, so that the number of collisions is 1.6×10^{28} per cc. per sec. Increase of molecular diameter increases the collision number, but as the former is generally associated with a larger molecular weight, the resulting effect will be small. Temperature also has relatively little effect, since only the square root is involved in (112); the number of collisions is, however, proportional to the square of the molecules in unit volume and hence to the square of the pressure.

Triple Collisions.—The number of occasions on which three molecules meet, called triple collisions, is of interest in the study of reaction velocities, but its value can only be estimated approximately. It may be supposed that if a third molecule approaches while a pair of molecules are still within a molecular distance of each other, the effect is equivalent to a triple collision. A reasonable consequence of this view, although it cannot be proved rigidly, is that the probability of a triple collision bears the same ratio to the chance of a dual collision as the molecular diameter bears to the mean free path in the gas under consideration. If this is so, then

$$Z_3/Z_2 \approx \sigma/l, \quad (113)$$

Z_2 and Z_3 being the numbers of dual and triple collisions, respectively, under the given conditions. Introducing the values of Z_2 and l , from (112) and (107), respectively, it follows that

$$Z_3 \approx 2\sqrt{2}\pi n^3\sigma^5(\pi RT/M)^{\frac{1}{2}} \quad (114)$$

for the number of triple collisions per sec. per cc. Since l is of the order of 10^{-5} cm. at 1 atm. pressure, whereas σ is about 10^{-8} , it is seen from (113) that the number of triple collisions is about 10^{-3} times the number of dual collisions at atmospheric pressure. At lower pressures the ratio is smaller for l increases whereas σ remains approximately constant.

Collisions with Walls.—By employing a method somewhat similar to that used in deriving the value for the pressure due to impacts on the

walls of the containing vessel (p. 250), it can be deduced that the number of molecules striking 1 sq. cm. of wall is on the average $\bar{c}n/4$ per sec., where n , as before, is the number of molecules in 1 cc. of the gas, and \bar{c} is their mean velocity. If m is the mass of each molecule, then the total mass of gas μ striking 1 sq. cm. of wall per sec., is given by

$$\mu = m\bar{c}n/4. \quad (115)$$

The quantity mn is the weight of unit volume, and so it may be put equal to M/V , where M is the molecular weight and V is the gram molecular volume; further, since $PV = RT$ for 1 mole, mn may be replaced by MP/RT , so that

$$\mu = \bar{c}MP/4RT. \quad (116)$$

Inserting the value for \bar{c} from (80), that is $(8RT/\pi M)^{1/2}$, it follows that

$$\mu = P(M/2\pi RT)^{1/2} \quad (117)$$

or

$$\mu = (P\rho/2\pi)^{1/2}, \quad (118)$$

where ρ , the density of the gas, is equal to M/V . Although the derivation of the mass of gas striking the surface in unit time has been considered from the standpoint of impact on the walls of a containing vessel, the result applies equally to impacts with any solid or liquid surface exposed to the gas.

It will be seen later (p. 443) that when a solid or liquid together with its vapor is placed in a closed vessel, an equilibrium is ultimately set up, at constant temperature, when the number of molecules leaving the surface of the solid or liquid is equal to the number returning from the vapor phase. The pressure exerted by the vapor is then the saturation vapor pressure of the liquid or solid. If the kinetic theory and the ideal gas laws may be assumed to hold for the vapor, assumptions that are approximate at least, then (117) and (118) give the rate at which molecules condense, provided every molecule striking the surface is retained*; they also give the rate at which they evaporate from 1 sq. cm. of surface. The value of P to be employed is, of course, the saturation vapor pressure at the given temperature. The vapor pressure of water at 15° c. is 1.28 cm. of mercury, and this is equivalent to $1.28 \times 981 \times 13.6$, i.e., 1.71×10^4 dynes per sq. cm.; inserting this in (117), and taking the molecular weight as 18 and R as 8.3×10^7 ergs deg.⁻¹ mole⁻¹, μ is found to be 0.156 g. Since one molecule of water weighs 3.01×10^{-23} g. (see p. 264), it follows that in every second approximately 5.2×10^{21} molecules leave, and the same number return to, each sq. cm. of water surface in equilibrium with its vapor at 15°.

Effusion.—The passage of a gas through a small orifice in a thin wall is known as **effusion**. If the diameter of the hole is small in comparison with the mean free path, then a molecule on reaching the opening is unlikely to collide with another but will pass right through; it follows

* For certain purposes a coefficient α is introduced to allow for the failure of every molecule to be held by the surface.

then that the number of molecules passing through the orifice is almost equal to the number which would normally strike an area of wall equal to that of the hole. The mass of gas effusing would thus be μ g. per sec. per sq. cm., where μ is given by (117) or (118); the corresponding volume is obtained on dividing by the density, i.e., μ/ρ cc., and so it follows from (118) that

$$\text{Rate of effusion in cc. per sec. per sq. cm.} = (P/2\pi\rho)^{\frac{1}{2}}. \quad (119)$$

At constant pressure, therefore, the rate of effusion, like that of diffusion, is inversely proportional to the square root of the density; this law, which should only hold at low pressures, when the mean free path is relatively large, was first discovered empirically by T. Graham (1829). It is doubtful whether in his experiments the condition that the size of the hole is small relative to the mean free path was satisfied, so that the gas streamed, rather than effused, through the hole. It appears, nevertheless, that the rate of streaming is also inversely proportional to the square root of the density of the gas.

It is seen from (119) that the density of a gas, at known pressure, could be evaluated by measuring the rate of effusion; alternatively, the densities of two gases could be compared by determining the time required for a definite volume of each gas to effuse, or stream, through a small hole at a given pressure. This principle was first employed by R. Bunsen (1857) for studying gas densities; it was also used by A. Debierne (1910) to determine the density, and hence the atomic weight, of radon. If the density of the gas or vapor is known, (118) or (119) may be utilized for the measurement of vapor pressures (see p. 449).⁴

Viscosity of Gases.—If two layers of a gas are considered, one of which is streamed over the other, then as a result of the interchange of molecules, on account of their continual movement, some of the momentum of the streaming layer will be transferred to the second layer. The net effect will be a decrease in the rate at which one layer moves with respect to the other. As a result of this internal friction, force must be continuously applied in order to maintain the movement. The retarding influence of a stationary layer of gas on a moving one is called **viscosity**. The **coefficient of viscosity** is defined as the force per unit area, expressed in dynes per sq. cm., required to maintain unit difference of velocity, i.e., 1 cm. per sec., between two parallel layers 1 cm. apart. The corresponding unit, i.e., 1 dyne sec. cm.⁻², is called a **poise**, in honor of J. L. M. Poiseuille (1844) who was a pioneer in the study of viscosity.

Imagine a gas flowing on a horizontal plane, and let q be the velocity gradient in a vertical direction, resulting from viscosity. Suppose the gas has zero velocity at the plane; then if the gradient is uniform the average velocity of flow at any level is given by the product of the gradient and the height of that level above the horizontal plane. Suppose l is the mean free path of the molecules; then, on the average, all molecules coming from a distance l cm. below a given level of height x will make their first colli-

sions, and hence their first transfer of momentum, when they reach the latter level. Since these molecules originate at a mean distance of $x - l$ above the horizontal plane over which the gas is moving, their average velocity of flow will be $(x - l)q$.

It may be supposed, as an approximation, that one-third of the gas molecules are moving in each of three directions at right angles. Consequently, one-sixth of the molecules will be moving upward and one-sixth downward, at any instant. If n is the number of molecules per cc., and \bar{c} is their average speed, $\frac{1}{3}n\bar{c}$ molecules will be moving upward, and an equal number downward, through an area of 1 sq. cm. in 1 sec. It follows, therefore, that the momentum transferred by the molecules moving upward per sq. cm. is $\frac{1}{3}mn\bar{c}(x - l)q$ per sec., where m is the mass of a molecule. Similarly, by considering the downward movement of the molecules from a distance l above the level x , it is seen that the downward transfer of momentum is $\frac{1}{3}mn\bar{c}(x + l)q$ per sq. cm. per sec. The net transfer of momentum per sec., i.e., the force acting, is thus the difference between these quantities, i.e., $\frac{1}{3}mn\bar{c}lq$, and for unit velocity gradient this becomes $\frac{1}{3}mn\bar{c}l$, per sq. cm. By the definition given above this is equal to the coefficient of viscosity, η ; thus,

$$\eta = \frac{1}{3}mn\bar{c}l. \quad (120)$$

Substituting the value for l given by (107), it follows that

$$\eta = \frac{1}{3}m\bar{c}/\sqrt{2\pi}\sigma^2. \quad (121)$$

Since the average velocity of a molecule is independent of the gas pressure (p. 268), it is seen that the viscosity coefficient likewise should not vary with pressure; the experimental verification of this deduction from the kinetic theory by A. Kundt and E. Warburg (1875) was regarded at the time as one of the strongest arguments in its favor.

The only term in (121) affected by temperature is the mean velocity, and so it follows that the coefficient of viscosity should be proportional to the square root of the absolute temperature. In actual practice it is found that it increases more rapidly than implied by this relationship. According to W. Sutherland (1893), the variation of viscosity with temperature can be represented by

$$\eta = \eta_0\sqrt{T}/(1 + C/T), \quad (122)$$

η_0 and C being constants for the gas; the latter is frequently known as Sutherland's constant. The deviation from ideal behavior appears to be due to the mutual attraction of the molecules, a factor which has been disregarded in the simple kinetic theory, but is of great importance as will be seen later (p. 286).

Equation (120) can be written in still another form by utilizing the fact that mn , the mass of unit volume, is equal to ρ , the density; thus,

$$\eta = \frac{1}{3}\rho\bar{c}l, \quad (123)$$

from which it is seen that a knowledge of viscosity, density and mean speed of the molecules of a gas permits the calculation of its mean free path. For example, the coefficient of viscosity of hydrogen at 0°C. is 8.41×10^{-5} poise, the density is 9×10^{-5} g. per cc., at 0° and 1 atm. pressure, and the mean velocity, calculated from (80), is 1.69×10^5 cm. per sec.; it follows from (123) that

$$l = \frac{3\eta}{\rho \bar{c}} = \frac{3 \times 8.41 \times 10^{-5}}{9 \times 10^{-5} \times 1.69 \times 10^5} = 1.66 \times 10^{-5} \text{ cm.},$$

which is of the same order as calculated from (107). It may be pointed out that the agreement between the values obtained in these two different ways provides evidence for the approximate accuracy, at least, of the Avogadro number as used in the calculation on p. 275.

Mean Free Path and Other Properties of Gases.—The heat conductivity of a gas is clearly related to the movement of the molecules, and application of the kinetic theory leads to the equation

$$k = \frac{1}{3} mn \bar{c} l c_v, \quad (124)$$

where k is the thermal conductivity, that is, the quantity of heat in ergs flowing per sec. through 1 sq. cm. under a temperature gradient of 1° per cm., and c_v is the specific heat of the gas at constant volume. Comparison of (120) and (124) indicates that k should be equal to ηc_v , although when other factors are taken into consideration, which are neglected in the simple kinetic theory treatment, it appears that the thermal conductivity will be somewhat greater; experimental measurements indicate that k is from 1.4 to 2.5 times ηc_v .

It has been seen from general considerations (p. 254) that the rate of diffusion of a gas is related to the speed of the molecules, and it was stated that the former is relatively slow because of molecular collisions; there should thus be a connection between the rate of diffusion, the speed of the molecules and their mean free path. Application of the kinetic theory leads to the equation

$$D = \frac{1}{3} \bar{c} l = \frac{1}{3} \bar{c} / \sqrt{2} n \sigma^2, \quad (125)$$

where D is the number of molecules of gas diffusing through 1 sq. cm. in 1 sec., when the concentration gradient is unity. For the diffusion of one gas or vapor into another, e.g., into air, the formula

$$D = \frac{\sqrt{(\bar{c}_1)^2 + (\bar{c}_2)^2}}{3\pi n \sigma_{1,2}^2 (1 + \alpha)} \quad (126)$$

has been derived; \bar{c}_1 and \bar{c}_2 are the mean velocities of the two kinds of molecules present, n is the total number of molecules per cc., $\sigma_{1,2}$ is the mean collision diameter of the two kinds of molecules, and α is a correction for the persistence of velocity, that is, the tendency for a molecule to continue to move in its initial direction, after collision.

It is seen that the mean free path of a gas may be evaluated from (123), (124) or (125) by making measurements of viscosity, thermal conductivity or diffusion. Further, if the value of n , the number of mole-

cules in unit volume, is known then it is possible to derive values of the collision diameters of gas molecules. Some of the results obtained in this manner are quoted in Table 32; they can be regarded as approximate only, because of the failure to make all the necessary corrections, but at least they give some indication of the magnitude of molecular diameters.

TABLE 32. MOLECULAR DIAMETERS FROM MEAN FREE PATHS

Substance	Viscosity	Heat Conductivity	Diffusion
Hydrogen	2.74 A.	2.72 A.	2.72 A.
Helium	2.18	2.20	—
Nitrogen	3.75	3.78	3.84
Oxygen	3.61	3.62	3.64
Carbon dioxide	4.60	4.82	4.38

The Loschmidt Number.—If the diameter of a molecule could be estimated in an independent manner, then it is clear that if the mean free path is known, the above calculation could be reversed so as to give an approximate value of the number of molecules in 1 cc. of gas. The first attempts in this direction were made by J. Loschmidt (1865), and the resulting figure, for 0° c. and 1 atm. pressure, is consequently called the **Loschmidt number**. The method used was based on the assumption that at low temperatures the volume of a liquid might be taken as approximately equal to the volume of the molecules considered as spheres * with diameters equal to their collision diameters. The effective volume of each molecule would then be $\frac{1}{6}\pi\sigma^3$, and hence 1 cc. of gas, containing n molecules at 0° c. and 1 atm., should occupy a volume of $\frac{1}{6}n\pi\sigma^3$ as liquid at low temperature; the normal density of the gas should thus be $\frac{1}{6}n\pi\sigma^3$ times the density of the liquid. If ρ_g and ρ_l represent the densities of gas and liquid, respectively, then

$$\rho_g/\rho_l = n\pi\sigma^3/6. \quad (127)$$

Further, since by equation (107) the mean free path can be expressed in terms of n and σ , it is possible from a knowledge of the gas and liquid densities, and the mean free path, obtained from viscosity or other measurements, to calculate both the Loschmidt number and the molecular diameter. Some examples of the results secured from more recent data are recorded in Table 33. The molecular diameters are somewhat

TABLE 33. CALCULATION OF LOSCHMIDT NUMBER AND COLLISION DIAMETERS

Gas	ρ_g/ρ_l	l	σ	n
Nitrogen	1.46×10^{-3}	9.5×10^{-6} cm.	12×10^{-8} cm.	1.7×10^{18}
Oxygen	1.15×10^{-3}	10.3×10^{-6}	10×10^{-8}	2.2×10^{18}
Carbon dioxide	2.54×10^{-3}	6.3×10^{-6}	14×10^{-8}	1.9×10^{18}

* The volume of close packed spheres is actually 74 per cent of the total volume occupied.

larger than those given in Table 32 because of the uncertainty involved in the assumption made above; this implies that there is no free space between the molecules in a liquid, and such is certainly not the case. Nevertheless the results are seen to be of the correct order of magnitude, as also is the value of the Loschmidt number, although it has been already noted that from the accepted Avogadro number the number of molecules per cc. at 0° c. and 1 atm. is about 2.7×10^{19} . Since (107) and (127) used in the calculation involve the product of n and σ , it is obvious that the employment of more correct, i.e., smaller, values for σ would give better results for the Loschmidt number. More accurate values can also be obtained by using the density of the solid instead of the liquid.

Viscosity and Molecular Diameters.—As a result of more precise treatment, making allowance for the distribution of velocities, the persistence of velocity after collision, and intermolecular attractions, (121) takes a form (S. Chapman, 1916) which may be simplified to

$$\eta = \frac{0.499m\bar{c}}{\sqrt{2}\pi\sigma^2(1 + C/T)}, \quad (128)$$

where C is the Sutherland constant. This equation has been used, chiefly by A. O. Rankine and C. J. Smith (1921 *et seq.*), to evaluate the collision diameters of a number of gaseous molecules. A convenient method for the determination of gas viscosity is to measure the rate of flow through a cylindrical capillary tube of known internal diameter, and to apply the Poiseuille equation (see p. 497),

$$\eta = \frac{\pi(P_1 - P_2)r^4t}{8VL}, \quad (129)$$

where t is the time taken for a volume V of the gas to pass through a capillary tube of length L and radius r , when P_1 is the pressure at one end and P_2 that at the other end of the tube. The volume of gas is assumed to be measured at the mean pressure. By using a standard apparatus, with a constant difference of pressure, all the factors in (129) except η and t may be kept constant, so that the viscosity will be directly proportional to the time of flow of the definite volume of gas. Hence, for two different gases A and B at the same temperature, provided the dimensions of the apparatus remain unchanged,

$$\eta_A/\eta_B = t_A/t_B. \quad (130)$$

If the coefficient of viscosity of any gas, e.g., dry air, has been determined accurately by a direct method, then this gas may be employed to calibrate the standard apparatus mentioned above.

An apparatus satisfying the conditions for which (130) is applicable was devised by A. O. Rankine (1910) and is shown diagrammatically in Fig. 49. It consists of a closed circuit formed by a bent capillary tube, about 0.2 mm. diam., on the left, and a wider tube, 0.3 cm. diam., on the right; the latter

contains a pellet of mercury about 4 cm. long. The apparatus is thoroughly cleaned, filled with gas at atmospheric pressure and inverted so as to bring the mercury pellet above the mark *A* on the tube when it is restored to its normal position. The whole is then placed in a constant temperature bath and the time taken for the mercury to fall between the marks *A* and *B* is observed. This gives the time required by the gas to pass through the capillary tube under a definite pressure gradient determined by the size, i.e., weight, of the mercury pellet and the length of the tube. The corresponding time, under exactly the same conditions, is then determined for dry air, the viscosity of which is known; application of (130) then gives the viscosity of the experimental gas. Allowance for the effect of capillary forces at the mercury surface can be made by repeating the observations with the same pellet split into two parts, so that the capillary effect is doubled. For the application of Chapman's equation (128) it is necessary to know the Sutherland constant, and this is obtained from viscosity measurements at a number of temperatures [equation (122)]; the mass *m* of a single molecule is evaluated by dividing the molecular weight by the accepted Avogadro number.*

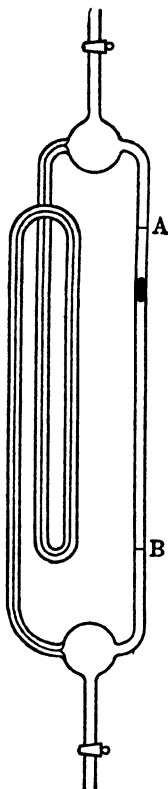


FIG. 49. Determination of the viscosity of a gas (Rankine)

The results in Table 34 show how (128) has been applied to determine the collision diameters of the hydrogen halides; the values are greater, as may be anticipated, than the true diameters obtained in other ways (see p. 569).

For vapors of condensable substances a modified form of apparatus is used; it consists of two vessels, at different temperatures containing the liquid, connected by a capillary tube at a temperature higher than either vessel. In this way a constant difference of pressure, represented by the difference of vapor pressure in the two vessels, is maintained at the ends of the tube through which the vapor flows, and the rate of flow may be measured by the increase in level of the liquid in the cooler vessel. The temperature of the latter should be low enough to ensure virtually complete condensation.

TABLE 34. COLLISION DIAMETERS FROM VISCOSITY MEASUREMENTS

Gas	Viscosity at 0° c.	Sutherland Constant	Collision Diameter
HCl	1.332×10^{-4}	357	2.86 Å.
HBr	1.710	375	3.16
HI	1.731	390	3.50

Measurement of Viscosity.—Although the flow method, two forms of which have been described above, for the determination of the viscosity of a gas is

* Rankine's experimental method is not regarded as accurate because of leakage of gas between the mercury and the glass.

convenient, it is not suitable for precision work, because of the numerous corrections which must be applied. The most accurate values for the viscosity of air, required for the evaluation of the electronic charge (p. 12), have been obtained by the **rotating cylinder method**. The apparatus consists of two concentric cylinders, separated by the gas under investigation. One of the cylinders is rotated at a constant speed, while the other is suspended by a torsion wire; the drag due to the rotating cylinder results in a definite deflection of the suspended one. From the rate of rotation and the angle of deflection, and other data, the viscosity of the gas may be calculated. A similar principle is used in the **oscillating disc method**. A flat horizontal disc is suspended from a torsion wire between two fixed discs, one above and one below. The suspended disc is set into oscillation, and the successive amplitudes are noted; from the damping of the oscillations it is possible to evaluate the viscosity of the gas in which the oscillations occur.⁵

Diffusion and Molecular Diameters.--The collision diameters of the molecules of solids and liquids with relatively low vapor pressures have been estimated from diffusion measurements (E. Mack, 1925) by the use of (126). The diffusion coefficient D was obtained by measuring the loss in weight of the solid or liquid, with known exposed area, when enclosed in a tube containing charcoal which absorbed the vapor as fast as the latter reached it. The following equation

$$D = \frac{wL}{At(\rho_1 - \rho_2)} \quad (131)$$

was then applied, where w is the weight of material lost by diffusion in time t , from an exposed area A when the distance between vaporizing surface and the charcoal is L ; ρ_1 and ρ_2 are the densities of the vapor at these two surfaces, respectively. The value of ρ_1 was calculated from the known vapor pressure of the substance, and ρ_2 was assumed to be zero since the charcoal is such a powerful absorbent. Some experimental results are given in Table 35; they are seen to be of the expected order.

TABLE 35. COLLISION DIAMETERS FROM DIFFUSION MEASUREMENTS

Substance	Diameter	Substance	Diameter
Benzene	5.7×10^{-8} cm.	Iodine	4.6×10^{-8} cm.
Naphthalene	7.4	Ammonia	2.9
Anthracene	8.0	Methane	3.2

MODIFICATION OF KINETIC THEORY FOR REAL GASES

The simple kinetic theory accounts for the properties of an ideal gas, but it requires some modification before it is applicable to real gases, since these, as already seen (p. 244), do not in actual practice obey the gas laws. There are, however, two assumptions made in the simple theoretical treatment which require correction; these are (i) that the molecules exert no attraction on each other, and (ii) that the volume of the molecules themselves is negligible in comparison with the total volume of gas.

Molecular Attraction.—The fact that gases can be converted into liquids, which possess the property of cohesion, implies that the molecules of a gas, as well as those of a liquid, attract one another. The result of this attraction, as explained below, is to make the pressure less than is to be expected theoretically. Direct proof of molecular attraction is to be found in the experiments of J. P. Joule and W. Thomson (see p. 190). A porous plug, of absorbent cotton or silk, was fixed in a boxwood tube and a stream of gas passed through; it was observed that at ordinary temperatures the gas was cooler after passing through the plug than before entering it. Later observers have worked with other porous materials, e.g., alundum and earthenware, and have confirmed the general observation that a gas cools itself on passing through a porous plug or throttle; this phenomenon is known as the **Joule-Thomson effect**. With most gases under ordinary conditions there is a fall of temperature, but hydrogen and helium exhibit an increase; if previously cooled, however, to low temperatures these substances behave normally. For an ideal gas, there should be no external work done in streaming from a higher to a lower pressure, and so the cooling observed is attributed to the work done in overcoming the attraction of the molecules for each other. Even in hydrogen and helium molecular attractions are present, but other factors tend to mask the effect except at low temperatures (see p. 293).

The Joule-Thomson Effect.—Before dealing with the corrections to be applied to the results of the kinetic theory, the Joule-Thomson effect will be considered briefly from the thermodynamic standpoint. The porous plug used in the experiment described above acts as a “throttle,” so that the pressure on each side remains practically constant throughout; let these pressures be P_1 , before entering the plug, and P_2 on the exit side. Suppose one mole of gas occupying volume V_1 at pressure P_1 and temperature T_1 streams through the plug, and let its volume be V_2 at the other end of the plug where its pressure is P_2 and temperature T_2 . The work done by the gas will be $P_2V_2 - P_1V_1$, since its volume is increased by V_2 at pressure P_2 while decreasing by V_1 at pressure P_1 . Since the whole process is assumed to be adiabatic, that is, no heat enters or leaves the system, it follows from the first law of thermodynamics that the work done by the gas must be equal to the decrease of internal energy (p. 198). If E_1 and E_2 are the internal energies before and after passage through the porous plug, then

$$P_2V_2 - P_1V_1 = E_1 - E_2; \quad (132)$$

$$\therefore E_1 + P_1V_1 = E_2 + P_2V_2. \quad (133)$$

By definition, however, the heat content H is equal to $E + PV$, and so it is clear that in the streaming process H remains unchanged, that is $\Delta H = 0$. The heat content is a definite property (p. 188), and its value depends only on the state of the system, e.g., on temperature and

pressure; hence dH is a complete differential, given by

$$dH = \left(\frac{\partial H}{\partial P} \right)_T dP + \left(\frac{\partial H}{\partial T} \right)_P dT. \quad (134)$$

In the Joule-Thomson experiment H is constant, i.e., dH is zero, and so it follows that

$$\left(\frac{\partial T}{\partial P} \right)_H = - \left(\frac{\partial H}{\partial P} \right)_T / \left(\frac{\partial H}{\partial T} \right)_P. \quad (135)$$

It was shown on p. 190 that $(\partial H / \partial T)_P$ is equal to C_P for 1 mole of gas, and utilizing (149), p. 236, in the form

$$\left(\frac{\partial H}{\partial P} \right)_T = V - T \left(\frac{\partial V}{\partial T} \right)_P, \quad (136)$$

(135) becomes

$$\left(\frac{\partial T}{\partial P} \right)_H = \frac{T \left(\frac{\partial V}{\partial T} \right)_P - V}{C_P}. \quad (137)$$

If $(\partial T / \partial P)_H$, often referred to as the **Joule-Thomson coefficient**, may be taken as constant over a small pressure range, (137) can be written as

$$\Delta T = \frac{T \left(\frac{\partial V}{\partial T} \right)_P - V}{C_P} \Delta P, \quad (138)$$

which is the equation for the differential Joule-Thomson effect, ΔT being the change (increase) of temperature for a change of pressure ΔP in the vicinity of the temperature T . In their experiments Joule and Thomson found that the fall of temperature was proportional to the difference in pressure on the two sides of the plug; this is in general agreement with (138).

For an ideal gas, $PV = RT$ and so $(\partial V / \partial T)_P$ is equal to R/P , i.e., to V/T , and it is at once evident that the numerator of (138) becomes zero; an ideal gas, therefore, should show no Joule-Thomson effect, in harmony with the thermodynamic definition (p. 190). For real gases the sign of the Joule-Thomson coefficient will depend on the relative values of $T(\partial V / \partial T)_P$ and V ; the connection between this and other properties of a gas will be considered shortly.

The Volume of the Molecules.—The fact that it has been necessary to postulate an appreciable collision diameter for the molecules of a gas, representing the distance of closest approach of the molecules upon collision, means that they must possess an effective volume. As a first approximation this may be taken as $\frac{4}{3}\pi\sigma^3$, where σ is the collision diameter, and hence it is of the order of 3×10^{-23} cc. per molecule. Since there are about 2.7×10^{19} molecules per cc. at ordinary temperature and pressure, the average volume *inhabited* by a molecule is 3.7×10^{-20} cc. Under normal conditions, therefore, the volume of the molecules is less

than 0.1 per cent of the total volume of the gas; at a pressure of about 100 atm., however, the molecules would inhabit a space only about twelve times their own volume, and the effect of the latter could not be neglected. Further, the volume of a gas can be reduced by pressure and cooling until the substance liquefies and then solidifies, but in the solid state there is considerable resistance to any further attempt at compression. It is apparent, therefore, that there is a limiting volume which may be taken as approximately equal to that of the molecules themselves. The effect of the finite size of the molecules will be to make the free space available for movement less than the actual measured volume of the gas. The number of impacts on the walls of the vessel, and hence the pressure, will consequently be greater than that calculated by the simple kinetic theory.

The existence of a more or less definite collision diameter means that there must be repulsive as well as attractive forces operating between molecules. The collision diameter then represents the distance between the centers of two molecules when the forces of repulsion become so great as to bring about a reversal in the direction of motion. It appears that the repulsive forces fall off extremely rapidly with increasing distance of separation of the molecules, much more rapidly than do the forces of attraction. The potential energy of a pair of molecules changes according to something like the inverse twelfth power of the distance on account of repulsion, but according to the inverse sixth power due to attraction (cf. p. 298). As a result, there is a net attraction between two molecules when they are an appreciable distance apart, and a net repulsion when they are very close together. As one molecule approaches another from a distance, the attractive force increases to some extent until a certain point when the repulsion begins to be detectable. The *net* attraction then diminishes as the intermolecular distance decreases, and eventually the repulsive force becomes dominant (see Fig. 87). The distance between the molecules when the rapidly increasing repulsive forces just balance the more slowly increasing attractive forces represents the so-called collision diameter.

The van der Waals Equation.—In order to modify the ideal gas equation so that it represents the behavior of real gases, allowance should be made for the forces of attraction and repulsion between the molecules. One of the first successful attempts was made by J. D. van der Waals (1873), following on the earlier work of G. A. Hirn (1867) and others. Consider a molecule in the interior of a gas; it is surrounded by other molecules equally distributed in all directions, and consequently they exert no resultant attractive force on the molecule under consideration. As the latter approaches the walls of the containing vessel, however, the uniform average distribution of the molecules changes to one in which gas molecules are present on one side only, so that a force is exerted tending to pull the molecule inward. It appears, therefore, that the instant any molecule is about to strike the walls of the containing vessel, and thus contribute its share toward the total gas pressure, the molecules in the

bulk of the gas exert a force having the effect of pulling the molecule away from the wall. The measured pressure P is thus *less* than the ideal pressure postulated by the simple kinetic theory, and a correction term P_a should be added to the former; thus the ideal pressure is equal to $P + P_a$. The force exerted on a single molecule about to strike the wall is evidently proportional to the number of molecules in the bulk of the gas, and consequently to its density (ρ). Further, the number of molecules striking the wall at any instant will also be proportional to the density of the gas, so that the total attractive force which is related to the correction term P_a is proportional to ρ^2 . If V is the volume occupied by 1 mole of gas, then ρ is proportional to $1/V$, and so it follows that

$$P_a \propto \rho^2 \propto \frac{1}{V^2}, \quad (139)$$

$$\therefore P + P_a = P + \frac{a}{V^2}, \quad (140)$$

where a is a proportionality constant; the corrected pressure is thus $P + a/V^2$. The term, a/V^2 , is a measure of the attractive force of the molecules; it is called the **cohesion pressure** or **internal pressure**, and is of importance in connection with the properties of liquids (see p. 479).

It has been seen above that the result of the finite size of molecules is to reduce the effective volume of the gas; in order to obtain the ideal volume, therefore, a correction term b , known as the **covolume**, must be subtracted from the measured volume V ; the corrected volume is thus $V - b$. It might be thought, on a first consideration, that b was itself equal to the volume of the molecules, but this is not the case; according to van der Waals the covolume is equal to four times the actual volume of the molecules, but other, slightly different, estimates have been made. A method of deriving the relationship between b and the volume of the molecules is given on p. 294.

The product of the *ideal* pressure and volume of 1 mole of a gas should be equal to RT , and so it follows, applying the corrections considered above, that

$$\left(P + \frac{a}{V^2} \right) (V - b) = RT. \quad (141)$$

This is called the **van der Waals equation**, and is one of the best known equations of state for real gases.

The values of a and b depend on the units used for expressing pressure and volume; if the pressure is in atm. and the volume in liters, then a is in liter²-atm. and b is in liters. Many of the results quoted in the literature and in tables of data are in Amagat units; in this method of expression the pressure is in atm., but the volume is in multiples of that at 0° c. and 1 atm. pressure, i.e., 22.4 liters. In order to convert a and b in Amagat units to liter-atm. values it is necessary to multiply the former by (22.4)² and the latter by 22.4. It was shown by van der Waals that a and b could

be calculated from the coefficients of thermal expansion and of compression of the gas, and insertion of the values in (141) gave better agreement with experiment than the simple gas equation. The correction terms a and b can, of course, be evaluated by inserting known values of P and V in the van der Waals equation and solving the resulting simultaneous equations. Alternatively, a and b can be calculated from the Joule-Thomson effect, as shown below, but most of the data recorded in the literature are evaluated from the critical constants of the gases (see p. 436). It will be seen later that a and b vary to some extent with temperature, and so the results quoted in Table 36 can only be regarded as approximate. It is

TABLE 36. VAN DER WAALS CONSTANTS IN LITER-ATM. UNITS PER MOLE

Gas	a	b	Gas	a	b
Hydrogen	0.245	2.67×10^{-2}	Hydrogen chloride	3.8	4.1×10^{-2}
Helium	0.034	2.36	Ammonia	4.0	3.6
Nitrogen	1.38	3.94	Ethylene	4.4	5.6
Oxygen	1.32	3.12	Chlorine	5.5	4.9
Carbon dioxide	3.60	4.28	Sulfur dioxide	6.7	5.6

evident that a increases with the ease of liquefaction of the gas; this is to be expected if it is a measure of the attraction between the molecules.

Low Pressures.—At extremely low pressures, when V is very large, both correction factors in the van der Waals equation may be neglected; the term a/V^2 is then very small and b is only a minute fraction of V , so that (141) becomes $PV = RT$. Under these conditions the gas obeys the ideal law, and so this may be regarded as representing the limiting behavior of gases at infinitesimally small pressures. At slightly greater pressures, it is generally possible to neglect b in comparison with V , so that the van der Waals equation becomes

$$\left(P + \frac{a}{V^2}\right)V = RT; \quad (142)$$

$$\therefore PV = RT - \frac{a}{V}, \quad (143)$$

so that PV is less than RT and decreases with increasing pressures, since V becomes less and hence a/V greater. The dip in the curve of PV against P (Figs. 40 and 41) is thus to be attributed to the a/V^2 , i.e., attraction, term in the van der Waals equation. This is in agreement with the statement made above that when the molecules are relatively far apart, as at low pressures, the attractive forces are dominant.

The van der Waals equation (141) may be written as

$$P(V - b) + \frac{a}{V^2}(V - b) = RT;$$

$$\therefore P(V - b) + a\left(\frac{1}{V} - \frac{b}{V^2}\right) = RT. \quad (144)$$

The difference between $1/V$ and b/V^2 is not large at fairly high pressures, and so the second term on the left-hand side of (144) may be neglected; hence,

$$\begin{aligned} P(V - b) &= RT, \\ PV &= RT + Pb, \end{aligned} \quad (145)$$

so that PV is now greater than RT , and increases with increasing pressure. The rise in the PV curve with increasing pressure, after passing through a minimum, is thus due to the b term. This is to be expected, since the repulsive forces become appreciable when the molecules are close together, as at high pressures. Since hydrogen and helium have exceptionally small values of a in comparison with b (Table 36) the influence of the latter predominates, and PV increases as the pressure is raised, even for small values of the latter, at ordinary temperatures.

Calculation of the Boyle Temperature.—The minima of the curves showing the variation of PV with P , e.g., Fig. 41, lie on a parabolic curve; as the temperature is raised a point is reached at which the minimum occurs, theoretically, at zero pressure, and at still higher temperatures no minimum is observed. Any gas will then behave in the same manner as hydrogen and helium do at ordinary temperatures. The condition for the PV minimum to occur when P is 0, i.e., for the Boyle point, T_B (p. 246), can be readily obtained from the van der Waals equation; for this purpose it may be written in the form

$$P = \frac{RT}{V - b} - \frac{a}{V^2}, \quad (146)$$

and multiplying through by V ,

$$PV = RT \left(\frac{V}{V - b} \right) - \frac{a}{V}. \quad (147)$$

Differentiating with respect to P at constant temperature, it follows that

$$\left(\frac{\partial(PV)}{\partial P} \right)_T = \left[\frac{RT}{V - b} - \frac{RTV}{(V - b)^2} + \frac{a}{V^2} \right] \left(\frac{\partial V}{\partial P} \right)_T. \quad (148)$$

The condition for a minimum in the curve of PV against P is that $\partial(PV)/\partial P$ should be zero; hence the expression in the square brackets may be equated to zero, thus giving

$$RT = \frac{a}{b} \left(\frac{V - b}{V} \right)^2. \quad (149)$$

If the minimum is to occur when P is 0, then V will be infinite, and the expression reduces to

$$\begin{aligned} RT_B &= \frac{a}{b}; \\ \therefore T_B &= \frac{a}{Rb}. \end{aligned} \quad (150)$$

At all temperatures above T_B the value of PV will always increase with increasing pressure, and it is evident that for hydrogen and helium this temperature will be relatively low since the a values are so small. For hydrogen, for example,

a is 0.245 and b is 2.67×10^{-2} in liter-atm. units, when R is 0.082; hence T_B can be calculated to be 112°K. , compared with the experimental value 106°K. On the other hand, for oxygen the data for a and b in Table 36 lead to a value of 529°K. for T_B ; this is not in such good agreement with the observed temperature, 423°K. , but it is nevertheless considerably higher than for hydrogen.

It was seen earlier that the lowest temperature above which PV increases continuously with increasing P was called the Boyle point (p. 246), because Boyle's law is then obeyed over a range of pressures; this can be shown to be so by means of the van der Waals equation. Writing (141) in the form

$$PV = RT - \frac{a}{V} + bP + \frac{ab}{V^2}, \quad (151)$$

obtained by multiplying out and rearranging the terms, it is evident that ab/V^2 may be neglected, since both a and b are small, provided the pressure is not too large. Further, V may be replaced by RT/P in the correction term a/V , so that

$$PV = RT - \frac{aP}{RT} + bP \quad (152)$$

$$= RT + P \left(b - \frac{a}{RT} \right). \quad (153)$$

If the temperature is the Boyle point, T_B , then it is equal to a/Rb , as seen above; if this is inserted in (153) it becomes

$$PV = RT_B, \quad (154)$$

so that the ideal gas laws are obeyed provided the pressure is not too high.

Calculation of the Joule-Thomson Effect.—If (152) is divided through by P , there is obtained

$$V = \frac{RT}{P} - \frac{a}{RT} + b \quad (155)$$

and differentiation with respect to temperature at constant pressure gives

$$\left(\frac{\partial V}{\partial T} \right)_P = \frac{R}{P} + \frac{a}{RT^2}. \quad (156)$$

Further, by rearrangement of (152),

$$RT = P(V - b) + \frac{aP}{RT}; \quad (157)$$

$$\therefore \frac{R}{P} = \frac{V - b}{T} + \frac{a}{RT^2}. \quad (158)$$

Substituting this value of R/P in (156), it follows that

$$\left(\frac{\partial V}{\partial T} \right)_P = \frac{V}{T} - \frac{b}{T} + \frac{2a}{RT^2}; \quad (159)$$

$$\therefore T \left(\frac{\partial V}{\partial T} \right)_P - V = \frac{2a}{RT} - b. \quad (160)$$

If this is substituted finally in (138), the value of the Joule-Thomson coefficient is seen to be given by

$$\frac{\Delta T}{\Delta P} = \frac{\left(\frac{2a}{RT} - b\right)}{C_P}. \quad (161)$$

By means of the values of a and b given in Table 36, the ratio of ΔT to ΔP may be calculated and compared with that observed for various gases. The results for 0°C. are recorded in Table 37; it should be noted that C_P must be expressed in the same units as a , b and R .^{*} The agree-

TABLE 37. THE JOULE-THOMSON COEFFICIENT

Gas	a (liter ² -atm.)	b (liters)	C_P (liter-atm.)	$\Delta T/\Delta P$ Calculated	$\Delta T/\Delta P$ Observed
Hydrogen	0.245	2.67×10^{-2}	0.29	-0.017	-0.03
Oxygen	1.32	3.12	0.30	+0.29	+0.31
Carbon dioxide	3.60	4.28	0.38	+0.73	+1.30

ment between the observed and calculated values of the Joule-Thomson coefficient is not exact, but the results are seen to be of the right order; the negative sign for hydrogen indicates that the temperature rises instead of falling, as with other gases at 0°C. , upon free expansion.

It will be noted from (161) that as long as $2a/RT$ is greater than b the right-hand side is positive; there is thus a fall of temperature as the gas streams through a porous plug from a higher to a lower pressure. If the temperature is such that $2a/RT$ is equal to b then ΔT will be zero, and there will be no Joule-Thomson effect. At higher temperatures $2a/RT$ is less than b and there is a rise of temperature when the gas passes through the plug. The Joule-Thomson inversion temperature T_1 where ΔT is zero occurs, therefore, when

$$\begin{aligned} \frac{2a}{RT_1} &= b; \\ \therefore T_1 &= \frac{2a}{Rb}. \end{aligned} \quad (162)$$

Comparison with (150) shows that the inversion temperature, as derived from the van der Waals equation, should be twice the Boyle point; the values for hydrogen and oxygen should then be 224°K. and 1058°K. , respectively. The observed inversion temperature for hydrogen is about 190°K. The low value may be attributed to the fact that the ratio of a to b for this gas is small; in other words, the attractive forces are relatively small, and hence the effect of the repulsive forces is dominant except at very low temperatures. The same is true for helium (see Table 36), which also has a negative Joule-Thomson coefficient at normal temperatures.

The simple treatment given above leads to the conclusion that there

^{*} The Joule-Thomson coefficient varies with the pressure; the observed values are for pressures of the order of 1 to 10 atm.

is only one inversion temperature and that it is independent of pressure; this is a result of the approximations made in the deduction, namely that the pressure is not too high and that ab/V^2 can be neglected. A more rigorous treatment based on the van der Waals equation, without the omission of any terms, gives the condition for the inversion temperature

$$\frac{2a}{RT_I} - \frac{3abP}{R^2T_I^2} - b = 0. \quad (163)$$

Since this is a quadratic there should be, as in fact has been observed, two inversion temperatures at each pressure; the plot of T_I against P gives a parabolic curve, and at pressures exceeding $a/3b^2$ the Joule-Thomson effect for the given gas is inverted at *all* temperatures. For oxygen this pressure should be about 450 atm., but the actual value is probably much higher.

The Volume of the Molecules.—The constant b of van der Waals's equation is related, as seen above, to the actual volume of the molecules. The relationship may be derived in several ways of varying complexity and exactness; the following is a relatively simple one. The volume v_m of a single molecule is equal to $\frac{4}{3}\pi(\sigma/2)^3$ where σ is the molecular collision diameter; since the centers of two molecules cannot approach closer than a molecular diameter (cf. Fig. 48, p. 274), however, the *effective radius* of a molecule may be taken as equal to its diameter, and so its *effective volume* is $\frac{4}{3}\pi\sigma^3$, which is $8v_m$. Suppose a space V is filled by n molecules of gas added one at a time; the volume available to the first molecule will be V , to the second it will be $V - 8v_m$, to the third $V - 2 \times 8v_m$, and so on until the n th molecule which will have an available volume

$$V - (n - 1)8v_m.$$

The average space available for each molecule can then be obtained by evaluating the geometric mean of these volumes; that is, by multiplying the n terms and extracting the n th root.

$$\begin{aligned} \text{Available space per molecule} &= \{V(V - 8v_m)(V - 2 \times 8v_m) \cdots (V - (n - 1)8v_m)\}^{1/n} \\ &= \{V^n - 8V^{n-1}v_m(1 + 2 + \cdots + n - 1)\}^{1/n} \\ &\approx V - 4nv_m. \end{aligned}$$

If the molecules themselves had no volume then the space available for each molecule would have been V , so $V - 4nv_m$ for a real gas replaces V for an ideal gas. It is clear, therefore, that $4nv_m$, which is four times the total volume of the molecules, is equal to b of van der Waals's equation; it is thus possible to calculate the molecular diameter from a knowledge of this quantity. The figures in Table 38 have been obtained in this manner from the data tabulated on p. 290; they may be compared with the results given in Table 32.

TABLE 38. MOLECULAR DIAMETERS FROM THE VAN DER WAALS CONSTANT

Substance	σ	Substance	σ
Hydrogen	2.76 A.	Carbon dioxide	3.24 A.
Helium	2.66	Ammonia	3.06
Nitrogen	3.14	Chlorine	3.40
Oxygen	2.92	Sulfur dioxide	3.55

The Clausius Equation.—Although the van der Waals equation gives a satisfactory qualitative representation of the behavior of a real gas, it fails in many quantitative aspects, as seen above (see also p. 437). One of the weaknesses of the treatment is that the quantities a and b are not independent of temperature; for example, if b is assumed to be constant, then a for carbon dioxide is found to decrease by about one-fifth between 0°C. and 100°C. , whereas if a is regarded as constant b increases in a similar proportion. Further, the effective volume of the molecules, as represented by b , apparently decreases as the pressure is raised. Many attempts have been made to improve the van der Waals equation and a few, which have attracted some attention, will be considered briefly. To allow for the variation of a with temperature, R. Clausius (1880) proposed the following equation, which involves the assumption that the molecular attraction factor is inversely proportional to the temperature, viz.,

$$\left[P + \frac{a}{T(V+c)^2} \right] (V-b) = RT. \quad (164)$$

For certain reasons, which will be explained in Chapter VI, an equation of state should not contain more than three arbitrary constants, e.g., a , b and R , and so it is necessary to assume that c is a function of one of the other constants (see p. 440). The introduction of a fourth constant enables better agreement with observed data to be obtained in many cases, but on the whole the advantages are not so great as to be compensated for by the difficulty of dealing mathematically with the equation; for this reason the Clausius equation has found relatively little application.

The Berthelot Equation.—D. Berthelot started by eliminating the fourth constant c from the Clausius equation, giving

$$\left(P + \frac{a}{TV^2} \right) (V-b) = RT, \quad (165)$$

which, on multiplying out, becomes

$$PV = RT + Pb - \frac{a}{TV} + \frac{ab}{TV^2}. \quad (166)$$

At not too high pressures the term ab/TV^2 , involving the product of two small quantities a and b , may be neglected; further if, as a first approximation, V in the term a/TV is replaced by RT/P , then (166) can be written

$$PV = RT + Pb - \frac{aP}{RT^2} \quad (167)$$

$$= RT \left(1 + \frac{Pb}{RT} - \frac{aP}{R^2T^3} \right). \quad (168)$$

The terms a , b and R are replaced by expressions involving the critical constants, based on experimental results (see p. 437); thus,

$$a = \frac{1}{3} P_c V_c^2 T_c, \quad b = \frac{1}{3} V_c, \quad R = \frac{8}{3} \frac{P_c V_c}{T_c},$$

with the result

$$PV = RT \left[1 + \frac{9}{128} \cdot \frac{PT_c}{P_c T} \left(1 - 6 \frac{T_c^2}{T^2} \right) \right]. \quad (169)$$

This is the Berthelot equation; it probably holds best at low pressures, and is applicable over a limited range only. Nevertheless, it has been found useful for applying corrections for deviations from ideal behavior.

By treating (167) in the same manner as the van der Waals equation on p. 292 and substituting the values for a , b and R given above, the following equation for the Joule-Thomson effect is obtained:

$$\Delta T = \frac{\frac{9}{128} \cdot \frac{RT_c}{P_c} \left(18 \frac{T_c^2}{T^2} - 1 \right)}{C_P} \Delta P. \quad (170)$$

The inversion point should occur when the temperature is equal to $\sqrt{18} T_c$.

The Dieterici Equation.—A modification of the van der Waals equation of a somewhat different type, is that proposed by C. Dieterici (1899). In the van der Waals deduction it is assumed that the density of the gas at the walls is the same as in the interior (p. 289). Since work has to be done against the attraction of the other molecules when bringing a molecule from the interior to the wall, it is apparent that the molecules at the wall will have a higher potential energy than those in the bulk of the gas. According to the classical distribution law, therefore, there will be fewer molecules per unit volume at the walls than in the interior; if ΔE is the excess potential energy per mole at the walls, then

$$\frac{n}{n_1} = e^{-\Delta E/RT}, \quad (171)$$

where n and n_1 are the numbers of molecules in unit volume at the walls and in the bulk of the gas, respectively. Since the pressure is proportional to the molecular concentration, it follows that

$$\frac{P}{P_1} = e^{-A/RT}, \quad (172)$$

where P is the observed gas pressure and P_1 is the pressure in the interior of the gas where the molecular attractions cancel each other; it is reasonable to suppose, therefore, that P_1 may be identified with the ideal pressure. The excess potential energy ΔE is replaced for convenience by the constant A in (172). If the correction for the volume of the molecules is the same as that used by van der Waals, then the product of ideal pressure and volume gives the result

$$\begin{aligned} P_1(V - b) &= RT; \\ \therefore P &= \frac{RT}{V - b} e^{-A/RT}. \end{aligned} \quad (173)$$

Although this equation is apparently sound theoretically it was not found satisfactory in practice, since A was not independent of pressure; an empirical correction was made, therefore, in which A was replaced by a/V , giving

$$P = \frac{RT}{V - b} e^{-a/RTV}, \quad (174)$$

which is **Dieterici's equation of state**.

At relatively low pressures, that is, when V is large, expansion of the exponential series gives

$$e^{-a/RTV} = 1 - \frac{a}{RTV};$$

$$\therefore P = \frac{RT}{V - b} - \frac{a}{V(V - b)}, \quad (175)$$

and since b is small in comparison with V , the denominator $V(V - b)$ may be replaced by V^2 , so that

$$P = \frac{RT}{V - b} - \frac{a}{V^2}.$$

This is, of course, the van der Waals equation; the Dieterici equation is consequently identical with the former at relatively low pressures.

If the methods described above for the calculation of the Boyle temperature (p. 291) and of the Joule-Thomson effect (p. 292) are applied to the Dieterici equation (174), the results for low pressures will be found to be identical, as is to be expected, with those derived from the van der Waals equation. It is, therefore, unnecessary to give the detailed deductions here.⁶

Other Equations of State.—Among equations proposed in recent years, mention may be made of that suggested by F. G. Keyes (1917) who used the concept of the nuclear atom as the basis of his deduction. He concluded that the van der Waals factor b was influenced by the surrounding molecules, and made allowance for this in the equation

$$P = \frac{RT}{V - \beta e^{-\alpha/V}} - \frac{A}{(V + l)^2}, \quad (176)$$

where A , α , β and l are constants, depending on the nature of the gas, which are determined from the experimental data. The equation has been applied to several gases and shown to represent their behavior over a range of temperature and pressure.

Various writers on the kinetic theory of gases, from the time of L. Boltzmann onward, have suggested that the deviation of gases from ideal behavior is to be attributed partly to molecular "aggregation"; this term covers both association involving valence or analogous forces, as well as that resulting from relatively prolonged contact between molecules on collision. The influence of this aggregation, which has the virtual effect of decreasing the number of

molecules in the gas, will be most marked at high pressures. Keyes has emphasized the fact that his equation can only be expected to apply when there is no aggregation, but if this occurs an additional factor, correcting for the diminution in the effective number of molecules, is required. The equation of J. A. Beattie and O. C. Bridgeman (1927) is apparently an attempt to include this correction; it takes the form

$$P = \frac{RT(1 - \epsilon)}{V^2} (V + B) - \frac{A}{V^2}, \quad (177)$$

where

$$A = A_0 \left(1 + \frac{a}{V} \right); \quad B = B_0 \left(1 - \frac{b}{V} \right); \quad \epsilon = \frac{c}{VT^3}.$$

The equation involves five empirical constants, A_0 , B_0 , a , b and c , in addition to R , but its authors claim that it is easy to manipulate, and that it has proved satisfactory for a considerable number of gases.⁷

General Equation of State.—The most general equation for the variation of pressure, volume and temperature of a gas is that of H. K. Onnes (1901); it is written

$$PV = RT \left[1 + \frac{B(T)}{V} + \frac{C(T)}{V^2} + \dots \right], \quad (178)$$

where $B(T)$, $C(T)$, etc., which are functions of the temperature, are called the second, third, etc., **virial coefficients**. This type of equation has proved useful for extrapolation purposes (cf. p. 313), and the equations of state already given can be expressed in the virial form if desired. The van der Waals equation, for example, is closely approximated by the expression

$$PV = RT \left[1 + \frac{1}{V} \left(b - \frac{a}{RT} \right) \right],$$

and hence according to this equation the second virial coefficient $B(T)$ is equivalent to $b - a/RT$. The values of these coefficients are determined essentially by the forces of interaction, i.e., attraction and repulsion, between the molecules, and in recent years attempts have been made to calculate them from a consideration of these forces.⁸

The Origin of Molecular Attraction.—It has been assumed hitherto, mainly as a result of the Joule-Thomson experiment and the existence of cohesion in liquids, that molecules attract one another, but the fundamental cause of this attraction has only become evident as a consequence of developments in wave mechanics. W. H. Keesom (1912, 1922) attempted to explain molecular attraction as due to the interaction of permanent dipoles (see p. 543) in the molecule. The magnitude and sign, i.e., whether it results in attraction or repulsion, of the interaction energy depends on the relative orientation of the dipoles, and hence the phenomenon is known as the **orientation effect**. If all orientations were equally possible, the resultant energy would be zero, but it was concluded from statistical considerations that certain orientations must be preferred. The mean interaction potential energy U_0 between two dipolar molecules is given by

$$U_0 = - \frac{3}{4} \mu^2 / r^3 kT, \quad (179)$$

provided the temperature is not too low; μ is the dipole moment (p. 543), r is the distance between the molecules, and k is the Boltzmann gas constant. The negative sign of the potential energy means that work has to be done on the molecules to separate them; that is to say, there is a resultant attraction. If the orientation effect were the only source of molecular interaction then the attraction would decrease rapidly with increasing temperature. This does not appear to be in accordance with the facts, and P. Debye (1920-21) showed that the moment induced by a dipolar molecule in an adjacent molecule would give rise to an additional attractive force, now known as the **induction effect**. The magnitude of the resulting energy, which is independent of temperature, is

$$U_I = -2\alpha\mu^2/r^6, \quad (180)$$

where α is the polarizability of the molecule (p. 534).

The suggestions outlined above meet, however, with two serious difficulties: in the first place, the interaction energies are applicable to a pair of molecules, but they are not additive for all the molecules in a gas or liquid. In fact, since a large number of different molecular arrangements are possible the energies may cancel out, and there will be no attractive force to account for cohesion. Secondly, many molecules which do not possess permanent dipoles, e.g., hydrogen, oxygen, methane, and the inert gases, exert appreciable attractions on each other; this difficulty was partly overcome by the suggestion that such molecules have "quadrupoles,"* but this is apparently impossible for the inert gas molecules.

A significant advance was made by F. London (1930) using the methods of quantum mechanics. It follows from the Heisenberg uncertainty principle that all molecules must possess energy even in their lowest states, i.e., the zero-point energy (p. 158), and consequently, even in monatomic molecules, the nuclei and electrons must be regarded as undergoing some kind of vibration with respect to each other. Although the physical significance probably cannot be exactly defined, it may be imagined that an *instantaneous* picture of a molecule would show various arrangements of nuclei and electrons having dipole moments. In a molecule with no permanent moment these rapidly varying dipoles when averaged over a large number of configurations give a resultant of zero. Nevertheless, the temporary dipoles are able to induce in other molecules dipoles in phase with themselves, so that as a result there is a net attraction between the molecules. As a first approximation the interaction energy has been calculated as

$$U_D = -\frac{3}{4}h\nu_0\alpha^2/r^6, \quad (181)$$

where h is the Planck constant and ν_0 is a characteristic frequency of the molecule. The forces arising in the manner described are said to be due to the **dispersion effect**, because the oscillations producing the attractive force are also responsible for the dispersion of light by the molecules (p. 533). An important difference between the dispersion effect and those previously mentioned is not only that the former is applicable to nondipolar molecules, but that it is additive for all the pairs of molecules in the gas, and so accounts for the phenomenon of cohesion.

* A quadrupole consists of a double charge of one sign with single charges of opposite sign at either side; the resultant *dipole* moment is zero.

The polarizability α of a molecule can be determined from its molecular refraction (p. 536), the dipole moment μ is generally available (p. 551), and so it is only necessary to know the frequency ν_0 in order that the three molecular interaction effects may be evaluated. When the optical dispersion of a gas has been studied experimentally ν_0 can be obtained from the dispersion equation, but London has shown that for many gases ν_0 is almost identical with ν_i , where $h\nu_i$, when expressed in electron-volts, is equal to the ionization potential. If the latter is known, therefore, a reasonably accurate value of ν_0 may be estimated without the necessity for measurements of refractivity. The relative magnitudes of the different types of molecular interaction for a number of polar and nonpolar molecules are given in Table 39. It is seen that the dis-

TABLE 39. RELATIVE MAGNITUDES OF MOLECULAR INTERACTION EFFECTS

Molecule	Dipole Moment	Orientation Effect	Induction Effect	Dispersion Effect
H ₂	—	—	—	11.3
A	—	—	—	57
N ₂	—	—	—	62
CH ₄	—	—	—	117
Cl ₂	—	—	—	461
CO	0.12 D.	0.0034	0.057	67
HCl	1.03	18.6	5.4	105
NH ₃	1.5	84	10	93
H ₂ O	1.8	190	10	47

persion effect alone is capable of producing considerable molecular interaction, and only for substances with relatively high dipole moments is the orientation effect of any great importance. The induction effect is always very small. For most simple molecules, therefore, the attractive forces, or van der Waals forces as they are frequently called, are almost exclusively determined by the dispersion effect. It should be noted that (179), (180) and (181) refer to the potential energy of interaction; the attractive force is obtained by differentiation with respect to r , so that it is inversely proportional to r^7 in each case.

The equation for the dispersion effect given above allows for the interaction between a dipole in one molecule with a dipole in another. The calculation has been extended to include dipole-quadrupole and quadrupole-quadrupole interactions; the former of these, represented by a term involving r^{-8} , may be appreciable, but the latter can be neglected.

It was seen earlier (p. 288) that there must be repulsive, as well as attractive, forces between molecules in order to account for such properties as the collision diameter and effective size. The forces of repulsion probably arise from the interaction of the electron fields as the molecules approach one another. According to the calculations of wave mechanics, it appears that the repulsive potential should be of the form $P(r)e^{-r/\rho}$, where $P(r)$ is a polynomial function of r , and ρ is a constant for the given molecular species. Replacement of $P(r)$ by a constant (cf. p. 409) yields a simpler expression, as a first approximation, but even this is not very amenable to mathematical treatment. For many purposes, therefore, an even simpler, more approximate, function Br^{-n} has been found convenient; the value of n is about 12 for a number of gases. A useful potential function, allowing for molecular attraction and repulsion, is therefore

$$U = -Ar^{-6} + Br^{-12}, \quad (182)$$

where the first term on the right-hand side represents the attractive energy and the second term the repulsive energy for various intermolecular separations. For simple molecules not possessing appreciable permanent dipole moments Ar^{-6} is almost identical with the dispersion effect, and so A is equal to $\frac{3}{2}h\nu_0\alpha^2$.

The so-called constants a and b of van der Waals are, of course, related to the attractive and repulsive forces acting between the molecules. The a term is determined mainly by the attractive forces, while b depends on both, since the effective size of the molecule is the result of a compensation between attraction and repulsion. By making certain simplifying assumptions, F. London (1930) and J. C. Slater and J. G. Kirkwood (1931) have been able to calculate the product of the two van der Waals constants, i.e., $a \times b$, from theoretical considerations; the results are in good agreement with the accepted values.⁹

THERMODYNAMICS OF IDEAL AND REAL GASES

The Fugacity.—It has been deduced (p. 238) for a system of constant composition that $(\partial F/\partial P)_T = V$. For 1 mole of a pure substance the free energy F and the chemical potential μ are identical, so that

$$\left(\frac{\partial \mu}{\partial P}\right)_T = V, \quad (183)$$

where V is now the molar volume. For an ideal gas, V is RT/P , so that integration of (183) gives

$$\mu = RT \ln P + \mu^0, \quad (184)$$

where the integration constant μ^0 depends only on the temperature and the nature of the gas. For a real gas (184) does not hold, and G. N. Lewis (1901) suggested the use of the **fugacity** (P^*) which may be defined by the equation

$$\mu = RT \ln P^* + \mu^0, \quad (185)$$

the ratio P^*/P , where P is the actual pressure, approaching unity as P tends to zero, i.e., as the gas approximates more closely to ideal behavior. For certain purposes it is convenient to define a quantity called the **activity** as the fugacity of a gas under any given conditions compared with the value in a **standard state**. For gases the latter is chosen as the ideal gas at unit pressure, that is to say, when its fugacity is unity. Adopting this standard, the activity of any gas is seen to be equal to its fugacity. Further, the **activity coefficient** may be expressed as the ratio of the activity to the pressure, i.e., P^*/P ; this is equal to unity for an ideal gas, and any departure from unity is a measure of the departure of the actual gas from ideal behavior.

To determine the fugacity of a gas the following procedure may be adopted: differentiate (185) with respect to pressure, at constant temperature, then

$$\left(\frac{\partial \mu}{\partial P}\right)_T = V = RT \left(\frac{\partial \ln P^*}{\partial P}\right)_T; \quad (186)$$

$$\therefore \left(\frac{\partial \ln P^*}{\partial P}\right)_T = \frac{V}{RT}, \quad (187)$$

or, at a definite temperature,

$$RTd \ln P^* = VdP, \quad (187a)$$

where V is the actual molar volume of the gas. For an ideal gas the volume of 1 mole is RT/P , and the quantity α , given by

$$\alpha = \frac{RT}{P} - V, \quad (188)$$

is a measure of the departure from ideal behavior, at a given temperature; from (188) there follows

$$\alpha dP = RT \frac{dP}{P} - VdP. \quad (189)$$

From this and (187a) it is seen that

$$RTd \ln P^* = RT \frac{dP}{P} - \alpha dP \quad (190)$$

and integration between the pressures 0 and P , recalling that P^*/P approaches unity at low pressures, gives

$$RT \ln P^* = RT \ln P - \int_0^P \alpha dP. \quad (191)$$

To evaluate the fugacity, therefore, it is necessary to plot α , as defined by (188), against the pressure, and to measure the area under the curve between pressures 0 and P . It is found experimentally that at low pressures α approaches a constant value; under these conditions (191) becomes

$$\ln \frac{P^*}{P} = - \frac{\alpha P}{RT}. \quad (192)$$

Further, at low pressures P^*/P is not very different from unity; it follows, therefore, that

$$\ln \frac{P^*}{P} \approx \frac{P^*}{P} - 1 = - \frac{\alpha P}{RT}; \quad (193)$$

$$\therefore \frac{P^*}{P} = 1 - \frac{\alpha P}{RT} = \frac{PV}{RT} \quad (194)$$

or

$$P^* = P^2V/RT, \quad (195)$$

where P and V are the observed pressure and volume of 1 mole of the gas. Equation (195) provides a simple method for evaluating fugacities at relatively low pressures.

MIXTURES OF GASES

Dalton's Law of Partial Pressures.—The connection between the total pressure of a mixture of gases and the pressures of the constituents was expressed by J. Dalton (1801) in the form of the **law of partial pressures** as follows: *the total pressure of a mixture of gases is equal to the sum of the partial pressures of the constituent gases.* The partial pressure is defined as the pressure each gas would exert if it *alone* occupied the volume

of the mixture at the same temperature. According to the kinetic theory a mixture of ideal gases must obey Dalton's law; since the molecules are assumed to have negligible volume and to exert no attraction on each other, the total pressure P will be given by

$$PV = \frac{1}{3}(N_1 m_1 \bar{c}_1^2 + N_2 m_2 \bar{c}_2^2 + \dots); \quad (196)$$

$$\therefore P = \frac{1}{3}N_1 m_1 \bar{c}_1^2 / V + \frac{1}{3}N_2 m_2 \bar{c}_2^2 / V + \dots, \quad (197)$$

where N_1 , m_1 , c_1 refer to one gas, N_2 , m_2 , c_2 to another, and so on (see p. 251). It follows also from the kinetic theory that each of the $\frac{1}{3}Nmc^2/V$ terms represents the pressure which the particular gas would exert at the same temperature if it alone occupied the whole volume V . In other words, each term in (197) represents the partial pressure of the gas indicated by the subscript; if these partial pressures are p_1 , p_2 , p_3 , etc., then

$$p_1 = \frac{1}{3}N_1 m_1 \bar{c}_1^2 / V; \quad p_2 = \frac{1}{3}N_2 m_2 \bar{c}_2^2 / V; \quad \text{etc.} \quad (198)$$

It follows, therefore, from (197) and (198) that

$$P = p_1 + p_2 + p_3 + \dots, \quad (199)$$

in agreement with Dalton's law. If the Maxwell relationship (p. 252) is introduced, that at constant temperature the mean kinetic energy per molecule, $\frac{1}{2}m\bar{c}^2$, is constant and equal to e_k , then from (198)

$$p_1 = \frac{2}{3}N_1 \left(\frac{e_k}{V} \right); \quad p_2 = \frac{2}{3}N_2 \left(\frac{e_k}{V} \right); \quad \text{etc.};$$

$$\therefore P = \frac{2}{3} \cdot \frac{e_k}{V} (N_1 + N_2 + N_3 + \dots);$$

$$\therefore p_1 = \frac{N_1}{N_1 + N_2 + N_3 + \dots} P = x_1 P, \quad (200)$$

$$p_2 = \frac{N_2}{N_1 + N_2 + N_3 + \dots} P = x_2 P, \quad (200a)$$

where x is the mole fraction* of each gas in the mixture. The partial pressure of a constituent in an ideal gas mixture is thus equal to the product of the total pressure (P) and the mole fraction (x) of that particular constituent.

The equation $PV = nRT$ applies to n moles of any ideal gas, and for a mixture of ideal gases, it follows that

$$PV = (n_1 + n_2 + n_3 + \dots)RT, \quad (201)$$

where P is the total pressure, V the total volume and n_1 , n_2 , n_3 , etc., the numbers of moles of the various constituents; the partial pressure of each

* The mole fraction of any constituent in any mixture (gaseous, liquid, or solid) is defined as the number of moles, or molecules, of that constituent divided by the total number of moles, or molecules, in the mixture.

gas is then given by

$$p_1V = n_1RT, \quad p_2V = n_2RT, \quad \text{etc.} \quad (202)$$

In the special case that the mixture contains a total of 1 mole, then $n_1 + n_2 + n_3 + \dots = 1$, and n_1 is identical with the mole fraction x_1 , and so on; hence from (202),

$$p_1 = x_1RT/V, \quad p_2 = x_2RT/V, \quad \text{etc.}, \quad (203)$$

where V is the volume of the 1 mole of gas mixture under the pressure P .

Thermodynamics of Gas Mixtures.—For any constituent of a mixture of ideal gases, it follows from (202) that $V = n_1RT/p_1$; the partial molar volume (\bar{V}_1) of this constituent is, therefore, at the given partial pressure p_1 (cf. p. 237),

$$\bar{V}_1 = \left(\frac{\partial V}{\partial n_1} \right)_{P, T, n_2, \dots} = \frac{RT}{p_1}, \quad (204)$$

assuming the temperature and the amounts of the other gases to remain constant. Further, as deduced on p. 239, with ∂p_1 replacing ∂P since the pressures of the other constituents can be assumed to be unchanged,

$$\left(\frac{\partial \mu_1}{\partial p_1} \right)_{T, n_1, n_2, \dots} = \bar{V}_1 = \frac{RT}{p_1}, \quad (205)$$

and hence, upon integration,

$$\mu_1 = RT \ln p_1 + \mu_P^0, \quad (206)$$

where μ_P^0 depends on the nature of the gas and its temperature. This relationship holds for any ideal gas, but for a real gas the partial pressure should be replaced by the theoretical fugacity p^* of the given constituent under the particular conditions, so that

$$\mu_1 = RT \ln p_1^* + \mu_P^0, \quad (207)$$

where p_1^*/p_1 is the activity coefficient which approaches unity as the total pressure tends to zero. The evaluation of p_1^* is not simple, but by a treatment analogous to that described on p. 302 an equation similar to (191) can be deduced.

Two further methods for expressing the chemical potential of the constituent of an ideal gas mixture have been found useful. At constant temperature and composition (205) can be put in the form

$$d\mu_1 = RT dp_1/p_1 = RT d \ln p_1, \quad (208)$$

and writing c_1 for n_1/V , that is, the concentration in moles per unit volume, e.g., 1 liter, p_1 can be replaced by RTc_1 , so that at constant temperature,

$$d\mu_1 = RT d \ln c_1 \quad (209)$$

and

$$\mu_1 = RT \ln c_1 + \mu^0. \quad (210)$$

Alternatively, as shown above, p_1 can be replaced by x_1P , where x_1 is the mole fraction of the given constituent and P is the total pressure, so that at constant

temperature and pressure

$$d\mu_1 = RTd \ln x_1 \quad (211)$$

and

$$\mu_1 = RT \ln x_1 + \mu_2^0, \quad (212)$$

where μ_2^0 depends on both temperature and pressure. For nonideal gases c_1 and x_1 should be replaced by the corresponding activity of the constituent in the mixture; it is related to the concentration or mole fraction just as the fugacity is related to pressure.

Real Gases and Dalton's Law.—As might be anticipated, real gases do not obey Dalton's law, which was enunciated on the basis of what would now be regarded as crude experiments. Unless the gases interact chemically, the total pressure is usually greater than the sum of the partial pressures, except for very highly compressed gases when there is strong intermolecular attraction. Some idea of the extent of the deviations from strict additivity will be obtained from the data in Table 40;

TABLE 40. POSITIVE DEVIATIONS FROM DALTON'S LAW

Total Pressure	Argon-Ethylene	Oxygen-Ethylene	Argon-Oxygen
30 atm.	0.85 atm.	— atm.	— atm.
40	1.6	—	—
50	2.7	2.5	0.5*
75	6.2	5.2	1.0*
100	8.0	7.2	1.45*

* Estimated values.

they are for mixtures containing approximately equimolar amounts of the two constituents in which the deviations are approximately maximal. Within the pressure range up to 100 atm. the deviations from Dalton's law evidently increase as the total pressure increases. At low pressures the deviations are quite small, and for values of the order of an atmosphere the law may be assumed to hold with fair accuracy. No serious error is, therefore, involved in the methods of gas analysis which depend on the additivity principle.

Various attempts have been made to formulate the additive rule in such a way as to be in better agreement with experiment at relatively high pressures. According to J. Willard Gibbs (1876) the pressure of a mixture of gases is equal to the sum of the pressures of the constituents when each exists by itself at the same temperature and with the same chemical potential as in the mixture. Since equality of temperature and potential implies a condition of equilibrium (p. 477), the total pressure should not be merely the sum of the separate pressures of the constituents, as required by Dalton's law, but the sum of the pressures of the individual gases which would be in equilibrium with the mixture (L. J. Gillespie, 1927), for example across a semipermeable membrane (p. 652). This rule has been tested by Gillespie in the following manner: the total pressure of ammonia, resulting from the dissociation of $\text{BaCl}_2 \cdot 8\text{NH}_3$, and of nitrogen

gas was measured manometrically. Since the dissociation pressure of the ammine is known at the experimental temperature, the value in equilibrium with the gas mixture is calculated from the thermodynamic equation relating dissociation or vapor pressure to the total pressure (p. 445). The equilibrium pressure of ammonia, for a given total pressure, is thus known and the result must be compared with that to be expected from the Gibbs rule. To do this, the total pressure P is written as the sum

$$P = f_{\text{NH}_3}(T, x_{\text{NH}_3}/V) + f_{\text{N}_2}(T, x_{\text{N}_2}/V),$$

where $p = f(T, x/V)$ represents an equation of state for each of the *pure* gases, x being the mole fraction of the gas in the mixture; the temperature T and the volume V are the same throughout. By successive approximations the value of V is found such that the two separate pressures add up to the observed total; the ammonia pressure obtained in this way should be equal to the experimental equilibrium pressure if the Gibbs additivity rule is correct. Although the agreement found was better than given by Dalton's ideal law, the discrepancies were still considerable.

G. N. Lewis and M. Randall (1923) suggested that a mixture of gases always formed an ideal solution (cf. p. 676), so that at any given temperature and pressure the fugacity of a constituent in the mixture, instead of its partial pressure, would be proportional to its mole fraction; thus,

$$p^* = xp_0^*, \quad (213)$$

where p_0^* is the fugacity of a constituent in the pure state at the pressure and temperature of the mixture, and p^* is the actual value in the mixture in which x is its mole fraction. The fugacity of each gas in a binary mixture can be calculated from a knowledge of the total volume as a function of pressure and composition, and the Lewis and Randall rule has been tested for ethylene-argon and hydrogen-nitrogen mixtures. Up to 25 atm. the results are satisfactory, but at higher pressures deviations are considerable; the rule is, however, a great improvement on Dalton's law.

Another possibility has been considered by E. P. Bartlett (1928), who suggested that at constant temperature the pressure exerted by one constituent of a gaseous mixture is equal to the product of its mole fraction and the pressure it would exert as a pure gas at a concentration equal to the total molar concentration of the mixture. Experimental tests prove this rule to be an improvement on that of Lewis and Randall. It has been claimed by J. R. Krichevsky (1937) that fugacities in excellent agreement with those observed in various gas mixtures may be obtained by squaring the values given by the Bartlett rule and dividing by those of the Lewis and Randall rule. There seems to be no theoretical justification for this purely empirical suggestion.¹⁰

The question of additivity in gas mixtures has been studied from another point of view; according to E. H. Amagat (1880, 1898) and A. Leduc (1896, 1898) volumes are more strictly additive than pressures. It is stated that the volume occupied by a mixture of gases is equal to the sum of the volumes the constituent gases would individually occupy at the temperature and pressure of the mixture. It is true that in many cases this rule is preferable to Dalton's law, e.g., for a mixture of nitrogen and hydrogen (E. P. Bartlett, 1928), but for other mixtures, e.g., ethylene

and argon (I. Masson, 1923), the deviations from additivity of pressure are less than those of volumes. A little consideration will show that for an ideal gas the Dalton and Amagat-Leduc rules are identical.

Even if two gases are not ideal, it is not impossible that their mixture should obey Dalton's law: in other words, their pressures might still be additive, were it not for the fact that in the mixture additional deviations are introduced because (i) the space in which the molecules of one gas are enclosed in the mixture is partly occupied by molecules of the other, and (ii) in addition to the attraction between molecules of the same kind, there will be an attraction between molecules of different kinds. These may be regarded as the essential causes of the departure from Dalton's law, and the rules given above are in a sense an attempt to allow for these factors.

If the composition of a gas mixture and its total pressure are known, the partial pressures of the constituent gases can be derived from (200) etc., only if the system behaves ideally. For practical purposes, however, it is convenient to define an effective partial pressure by (200), i.e., $p_i = x_i P$, where x_i is the mole fraction of any constituent and P is the total pressure, for any gas mixture. Dalton's law is then always applicable in the sense that the total pressure is equal to the sum of the defined effective partial pressures.¹¹

GAS DENSITIES AND MOLECULAR WEIGHTS

Normal Density of a Gas.—The **normal density** of a gas or vapor is defined as the weight of a **normal liter**, that is the weight of 1 liter of gas measured at 0° C., and at a pressure of 76 cm. of mercury at sea level in latitude 45°; these conditions are known as the standard temperature and pressure, and are abbreviated to s.t.p.* If the volume is not measured exactly at s.t.p., then the necessary adjustment must be made. Since the gas may show appreciable departure from ideal behavior, it is not justifiable, except for approximate purposes, to make the simple corrections based on the ideal gas laws (p. 192). If accurate results are required the normal density should be determined and corrections for deviations from ideality applied by one of the methods to be described shortly; this gives what is called the **limiting density** (ρ_0). It has been found by extrapolation to such conditions that Avogadro's law is applicable, that 1 mole of an ideal gas at s.t.p. should occupy 22.414 liters (see p. 193) so that the normal density of an ideal gas of molecular weight M should be $M/22.414$. This should be equal to the limiting density; hence,

$$\rho_0 = \frac{M}{22.414} \quad \text{or} \quad M = 22.414\rho_0. \quad (214)$$

* The abbreviation s.t.p. is frequently employed to indicate a temperature of 0° C. and a pressure of 1 atm., without stipulating the latitude or height above sea level.

Molecular weights determined in this manner have been utilized for the evaluation of accurate atomic weights.

For approximate work the gas laws may be assumed to hold, and the ideal equation employed; for 1 mole of gas this is $PV = RT$, and so for n moles, it takes the form

$$PV = nRT. \quad (215)$$

If V is the volume of w g. of gas, or vapor, of molecular weight M at pressure P and temperature T , then n in (215) is equal to w/M , so that

$$PV = \frac{w}{M} RT, \quad \text{or} \quad M = \frac{wRT}{PV}. \quad (216)$$

The value of R is known (p. 193) from measurements with gases of known molecular weight, and so M may be evaluated by determining the volume of a given mass of the gas or vapor, or the mass of a given volume, at known temperature and pressure; the molecular weight is thus obtained from what is virtually a measurement of density.

Measurement of Gas Density.—(a) The “globe” or “balloon” method was first employed for accurate work by H. V. Regnault (1845). A clean glass globe of known capacity is dried, evacuated and weighed; it is then filled with the pure dry gas at known temperature and pressure and re-weighed. In this way the weight of gas filling the globe is obtained. Errors due to buoyancy or to changes in the amount of moisture on the glass surface, resulting from alterations in the external conditions, are avoided by the use of a similar globe as a counterpoise. A small buoyancy correction is required, however, because the volume of the evacuated globe differs somewhat from that when filled with gas. For the determination of the normal density, measurements are made at 0° c. and at a pressure as near 1 atm. as possible, so that the necessary correction is small. There still remains adsorption of the gas (see p. 1194) on the interior surface of the globe as a possible source of error; this is small and generally ignored if concordant results are obtained after successive fillings and evacuations. In the earliest work with the Regnault method the globes employed were of about 10 liters capacity, but with improvements in the technique of weighing the size has been decreased to less than half a liter; buoyancy and other errors are thus considerably reduced.

(b) In one form of the “volumeter” method (E. W. Morley, 1895) a known weight of the gas is liberated, from a suitable generating apparatus which is weighed at the beginning and end of the experiment, and collected in a globe or globes of known volume at a definite temperature, e.g., 0° c. By means of a manometer attached to the globe, the pressure of the gas, whose weight, volume and temperature are known, can be determined. P. A. Guye (1904) has used a modified procedure: the globes are filled with gas at 0° and the pressure measured; the gas is then pumped out and absorbed in a weighed tube containing an appropriate reagent. The gain in weight of the latter gives the weight of the gas, and the difference between the initial and final pressures in the apparatus gives its pressure. If the gas is readily liquefied, e.g., ammonia, it may be compressed into a strong tube which can be detached from the apparatus and weighed (O. Maass, 1918). The chief advantage of the volumeter method

is that large volumes of gas may be employed, but the weight is determined in a small piece of apparatus, e.g., the generating vessel, an absorption bulb, or a condensation tube.¹²

(c) In recent years the "buoyancy" method, involving the use of a form of microbalance, has received much attention; it was adopted by R. Whytlaw-Gray and W. Ramsay (1910) to determine the density of radon, because only small quantities of gas are required, but latterly the method has attracted interest because of the accuracy of which it is capable. The form of microbalance used by R. Whytlaw-Gray and his collaborators (1931 *et seq.*) is shown diagrammatically, in plan, in Fig. 50; it is made entirely of quartz and consists of an

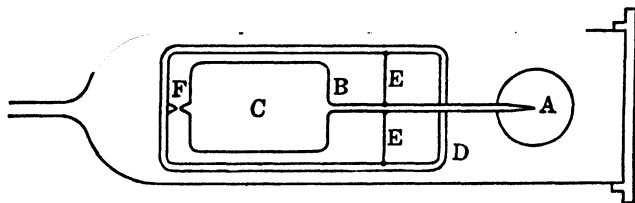


FIG. 50. Buoyancy microbalance (Whytlaw-Gray)

evacuated buoyancy bulb *A* of about 8 cc. capacity, suspended by a hook and a very thin quartz fiber from one end of the beam *B*, the other end of which is fixed to the plate *C* whose area is equal to that of the bulb. The beam is attached to the framework *D* of the balance by two fine tightly-stretched quartz fibers *EE*, about which it can swing. The whole is enclosed in a tube of about 2 to 3 cm. diameter and 8 to 10 cm. length, the temperature of which can be carefully controlled and which can be filled with gas at a measured pressure. The tube is first evacuated and then a gas is slowly admitted. The buoyancy effect causes the bulb *A* to rise until the pointer at the end of *C* is exactly on a level with that (*F*) on the framework; the pressure of the gas is then noted. The apparatus is again evacuated and the procedure repeated with another gas; if P_1 and P_2 are the pressures recorded, then the two gases must have equal densities at these pressures. For an ideal gas $PV = wRT/M$, and since w/V is equal to the density ρ ,

$$P = \rho RT/M. \quad (217)$$

For two gases, at constant temperature, having the same density at two pressures P_1 and P_2 , it follows that

$$P_1/P_2 = M_2/M_1. \quad (218)$$

This equation applies only if the gases are ideal; the relationship for real gases is given below.¹³

Corrections for Departure from Ideal Behavior.—The fundamental principle involved in the correction for failure to obey Boyle's law is the assumption, for which there is experimental evidence (p. 247), that the lower the pressure the smaller the deviations from ideal behavior; hence at infinitely small pressures all gases should behave ideally (D. Berthelot, 1898; P. A. Guye, 1908). Let V_1 be the volume of 1 mole of gas at pressure P_1 , equal to 1 atm., and temperature 0°C. , and V_0 that at an in-

finitely small pressure P_0 , when the ideal gas laws are obeyed; then a quantity λ may be defined by

$$P_0 V_0 / P_1 V_1 = 1 + \lambda. \quad (219)$$

If ρ_n is the normal density, that is, the weight of 1 liter at 0°C . and 1 standard atm. pressure, then V_1 , the volume in liters of 1 mole at S.T.P., is equal to M/ρ_n , where M is the molecular weight of the gas; substitution in (219) gives

$$M = \frac{P_0 V_0}{P_1} \cdot \frac{\rho_n}{1 + \lambda}. \quad (220)$$

As the gas behaves ideally at the very low pressure P_0 , the product $P_0 V_0$ may be replaced by RT , which is equal to 22.414 liter-atm. units at 0°C .; this is the correct unit as volumes are expressed in liters. Since P_1 is 1 atm., it may be put equal to unity, so that (220) becomes

$$M = 22.414 \rho_n / (1 + \lambda). \quad (221)$$

Comparison with (214) shows that $\rho_n/(1 + \lambda)$ is equal to ρ_0 , the limiting density of the gas, in g. per liter, corrected for deviations from the ideal gas laws. In the derivation of (221) the only assumption made is that at infinitely small pressures PV is equal to RT , and so it may be regarded as exact. The determination of λ would thus provide an accurate method for the evaluation of molecular weights from gas densities. Instead of utilizing the value of R based on other measurements, the molecular weights of two gases 1 and 2 may be compared by the relationship

$$\frac{M_1}{M_2} = \frac{\rho_1(1 + \lambda_2)}{\rho_2(1 + \lambda_1)}, \quad (222)$$

derived from (221).

It has been found experimentally that for all gases except the most compressible, the product PV is almost exactly a linear function of P at pressures up to 1 atm.; assuming the relationship continues down to extremely low pressures, it is possible to write

$$PV = P_0 V_0 (1 - AP), \quad (223)$$

where P_0 and V_0 refer to such conditions that the gas behaves ideally. The constant A , called the **compressibility coefficient**, may be evaluated from this equation without a knowledge of $P_0 V_0$, by measuring PV for at least two different values of the pressure. If P is 1 atm. and the temperature is 0°C ., then PV may be replaced by $P_1 V_1$, having the same significance as above, for 1 mole of gas; (223) then becomes

$$P_1 V_1 = P_0 V_0 (1 - A), \quad (224)$$

P in the parentheses being put equal to unity, so that the pressures are in atm. Comparison of (224) and (219) shows that provided the plot of PV

against P is linear, $1 - A$ is equivalent to $1/(1 + \lambda)$, so that (221) may be written

$$M = 22.414\rho_n(1 - A). \quad (225)$$

The limiting density, and hence the molecular weight, may thus be determined provided a number of measurements of the variation of PV with pressure have been made, so that A may be evaluated. The equation

$$\frac{M_1}{M_2} = \frac{\rho_1(1 - A_1)}{\rho_2(1 - A_2)} \quad (226)$$

is sometimes used for the comparison of molecular weights of two gases; the measurements of density and compressibility are made under similar conditions, so that unexpected errors tend to cancel out. The results in Table 41 may be taken to illustrate the application of (226); the second gas is oxygen in each case, so that M_2 is 32.000, A_2 is 0.0006, and ρ_2 is 1.4290 g. per liter.

TABLE 41. COMPRESSIBILITY CORRECTION FOR DEPARTURE FROM IDEALITY

Gas	Normal Density	A	Molecular Wt.
Ethylene	1.2606	0.0072	28.051
Carbon dioxide	1.9767	0.0066	44.010
Nitrous oxide	1.9777	0.0070	44.016

For fairly compressible gases PV does not vary in a linear manner with P , and so the foregoing method for evaluating $1 + \lambda$ cannot be used. One alternative is to evaluate PV/P_1V_1 , where P_1 is 1 atm. and P is a lower pressure, for several values of P , and then to extrapolate graphically to zero pressure. This gives P_0V_0/P_1V_1 , which is equal to $1 + \lambda$ (G. P. Baxter, 1922).

A modification of the procedure just described is to plot ρ/P , where ρ is the density at pressure P , for various pressures below atmospheric, and to extrapolate to zero pressure (P. A. Guye, 1905). The ideal gas equation (216) may be written in the form

$$M = \frac{\rho}{P} RT, \quad (227)$$

where ρ , the density of the gas, is substituted for w/V . This result will not apply to real gases at ordinary pressures, but if the extrapolated value of the quantity ρ/P , i.e., $(\rho/P)_0$, is used in (227), the equation

$$M = \left(\frac{\rho}{P} \right)_0 RT \quad (228)$$

will be valid for any gas. In this manner the molecular weight of the gas may be determined. In some instances the plot of ρ/P against P

is almost linear (cf. Fig. 51), but even for highly compressible gases, when the departure from a straight line is appreciable, it is not a difficult matter to obtain a reasonably accurate value of $(\rho/P)_0$ by graphical extrapolation. The treatment may be adapted, in the usual manner, for the comparison of the molecular weights of two gases, thus avoiding the use of the gas constant.

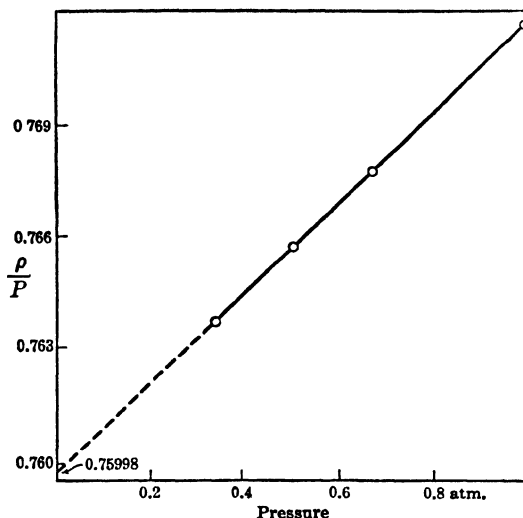


FIG. 51. Extrapolation of density of ammonia gas

In the limiting pressure method (W. Cawood and H. S. Patterson, 1936) use is made of the results obtained by the microbalance. Replacing P_0V_0 by RT for 1 mole of gas, (223) can be written as

$$PV = \frac{w}{M} RT(1 - AP) \quad (229)$$

for a weight w of gas; replacing w/V by the density ρ , it follows that

$$P = \frac{\rho}{M} RT(1 - AP). \quad (230)$$

if P_1 and P_2 are the pressures, as determined by the microbalance, at which two gases have the same density, then

$$\frac{P_1}{P_2} = \frac{M_2(1 - A_1P_1)}{M_1(1 - A_2P_2)}. \quad (231)$$

If the pressures P_1 and P_2 are very small, the compressibility corrections are negligible and the ratio of the two pressures becomes equal to the inverse ratio of their molecular weights [cf. equation (218)]. The ratio of the pressures is altered by changing the balance point of the micro-

balance, and linear extrapolation of the results to zero value of P_1 gives the ratio at an infinitesimally small pressure. The molecular weights of the two gases can thus be compared. The method may be illustrated by the results in Table 42; the reference gas is oxygen ($M_1 = 32.000$), and

TABLE 42. CALCULATION OF MOLECULAR WEIGHTS BY THE LIMITING PRESSURE METHOD

Ethylene		Carbon Dioxide		Nitrous Oxide	
P_1	P_1/P_2	P_1	P_1/P_2	P_1	P_1/P_2
428.1 mm.	0.879963	418.3 mm.	1.37764	418.6 mm.	1.37794
234.9 mm.	0.878507	234.5 mm.	1.37662	229.1 mm.	1.37680
0 mm.	0.876735	0 mm.	1.37532	0 mm.	1.37542
$M_2 = 28.056$		$M_2 = 44.010$		$M_2 = 44.013$	

the values of P_1/P_2 are extrapolated to zero pressure (P_1) of this gas. The agreement between the molecular weights obtained by the micro-balance method and those recorded above, based on direct measurements of density and compressibility, is very striking. It should be noted that the limiting density, and equivalent, methods described here are regarded as giving molecular weights of such accuracy that they can be employed for the purpose of determining atomic weights. The accepted values for carbon, 12.01, and nitrogen, 14.008, are based partly on data obtained in this way.

Methods Using Equations of State.—If the pressure-volume product can be expressed accurately by an equation of the form

$$PV = RT + \alpha/V + \beta/V^2 + \gamma/V^3 + \dots, \quad (232)$$

then λ can be determined by the equation

$$1 + \lambda = \frac{P_0 V_0}{P_1 V_1} = \frac{RT_0}{RT_0 + \alpha/V_1 + \beta/V_1^2 + \gamma/V_1^3 + \dots}, \quad (233)$$

where T_0 refers to 0°C. , i.e., RT_0 is 22.414 liter-atm. and V_1 is the volume of 1 mole at s.t.p. The value of V_1 is obtained by successive approximations, assuming in the first place that it is equal to 22.41, the constants α , β , γ , etc., being known from compressibility measurements. J. A. Beattie and O. C. Bridgeman (1928) have adapted their equation of state (p. 298) for the purpose of determining limiting densities. Since it involves five adjustable constants the equation should reproduce the behavior of a gas very closely, and accurate molecular weights should be obtained. Other limiting density methods involving equations of state have been proposed, but the results are probably not so reliable as those already recorded.

It has been seen that for 1 mole of gas $P_0 V_0$ may be put equal to RT , so that (219) becomes

$$P_1 V_1 = RT/(1 + \lambda). \quad (234)$$

Comparison with the Berthelot equation of state (p. 295) shows that

$$\frac{1}{1 + \lambda} = 1 + \frac{9}{128} \frac{P}{P_c} \cdot \frac{T_c}{T} \left(1 - 6 \frac{T_c^2}{T^2} \right). \quad (235)$$

The factor for converting normal to limiting densities can thus be derived from critical data; for example, for carbon dioxide P_c is 73.00 atm., T_c is 304.25° K., so that $1/(1 + \lambda)$ is found to be 0.9931. As the normal density of the gas is 1.9767, the molecular weight is 44.00.

Reduction of Critical Constants.—The following procedure, sometimes known as the method of reduction of critical constants, was suggested by P. A. Guye (1905). By taking P and V out of the two factors in the van der Waals equation of state, it becomes

$$\left(1 + \frac{a}{PV^2}\right)\left(1 - \frac{b}{V}\right) = \frac{RT}{PV} = \frac{P_0V_0}{PV}, \quad (236)$$

since, on the basis of the assumptions already made, RT may be equated to P_0V_0 . If the pressure is P_1 , i.e., 1 atm., and the volumes are expressed in terms of the normal volume V_1 , then (236) can be written

$$(1 + a)(1 - b) = P_0V_0/P_1V_1, \quad (237)$$

where P and V in the parentheses of (236) have each been given their value of unity. It follows, therefore, that

$$\frac{P_0V_0}{P_1V_1} = 1 + \lambda = (1 + a)(1 - b) \quad (238)$$

and hence, from (221),

$$M = \frac{22.414 \rho_n}{(1 + a)(1 - b)}. \quad (239)$$

The quantities a and b vary with temperature, and as those tabulated are almost invariably determined from critical data (see p. 436), it may be assumed that they apply at the critical temperature. Guye therefore proposed empirical equations to obtain the values at 0° C. as required for the application of (239). It is this uncertainty in a and b that makes the method less accurate than those already described. Nevertheless, very satisfactory results may be obtained; for carbon dioxide, it was estimated that a is 847×10^{-5} and b is 161×10^{-5} at 0°, the pressure being in atm. and the volume in terms of the normal volume. Taking the weight of a normal liter as 1.9767 g., the molecular weight of carbon dioxide is found to be 44.01; this is virtually identical with the accurate results in Tables 41 and 42.¹⁴

VAPOR DENSITY MEASUREMENTS

Approximate Molecular Weights.—The vapors of substances which are solid or liquid at ordinary temperatures deviate so considerably from ideal behavior that no great accuracy could be expected in the molecular weights without the expenditure of much labor. It is nevertheless of value to have means whereby approximate molecular weights may be ascertained relatively quickly, and a number of methods are available. They are generally referred to as **vapor density methods**, because they involve measurement of the weight and volume of the vapor. Since the determination of molecular weights is the primary object of the experi-

ment, the results are generally calculated by equation (216), on p. 308, viz.,

$$M = wRT/PV.$$

If the pressure is stated in atm. and the volume in liters, R is 0.082, so that

$$M = 0.082wT/PV. \quad (240)$$

The various experimental methods proposed involve different means for the determination of the weight w , volume V , pressure P and temperature T of the vapor; the more important of these will be described briefly.

Method of J. B. A. Dumas (1826).—The substance to be studied is introduced into a weighed glass bulb, about 250 cc. capacity, having a narrow outlet tube that can either be sealed off or closed by a stopcock. The apparatus is then heated in a constant temperature bath at about 20° above the boiling point of the experimental material. The heating is continued until the whole of the liquid has been converted into vapor and the air completely expelled from the bulb. The latter is now sealed up or closed, removed from the heating bath, cooled and weighed. As the weight of the bulb is known, the weight of the vapor filling it can be obtained, an allowance being made for the air contained in the "empty" bulb. The volume of the bulb, the temperature of the heating bath and the atmospheric pressure, provide, in addition, all the data necessary for determining the molecular weight. For the Dumas method at high temperatures porcelain and platinum globes have been employed.

Method of A. W. Hofmann (1868).—This method is based on an older one used by J. L. Gay-Lussac (1811). The apparatus consists of a graduated barometer tube containing mercury, surrounded by a jacket for maintaining a definite temperature by means of the vapor or a boiling liquid. A weighed quantity, e.g., 0.1 to 0.3 g., of the experimental substance is enclosed in a small loosely-stoppered glass tube, sometimes called a Hofmann bottle, and introduced into the apparatus by allowing it to rise through the mercury. The liquid vaporizes in the space above the mercury so that the level of the latter falls; this lowering gives the pressure of the vapor, and its volume is read off from the graduations on the tube. Since the temperature is known, the molecular weight can be calculated. An important advantage of the Hofmann method is that vaporization occurs under reduced pressure, so that the temperature may be well below the normal boiling point; substances which decompose on heating at atmospheric pressure may thus be studied. Further, as the volume of vapor is measured at a low pressure, only a small quantity of material is required. In a modified apparatus, used by W. Ramsay and S. Young (1887), the pressure in the barometer tube, as well as that over the boiling liquid for the heating jacket, can be varied; it is thus possible to make a series of measurements at different temperatures and pressures. On account of the volatility of mercury the Hofmann method cannot be used for temperatures over 250° c., and at 100° c. or above a correction should be applied for the vapor pressure of mercury.

Method of Victor Meyer (1878).—The method now to be described has probably attracted more attention than any other because of its simplicity and adaptability over a wide range of temperature. In its simplest form the apparatus consists of a tube about 75 cm. in length and 1 cm. in diameter widening

out to an elongated bulb at its lower end, and fitted with a side-tube near the top. From the side-tube downward it is surrounded by a jacket maintained at constant temperature by vapor from a boiling liquid; the temperature should be 20° to 30° above the boiling point of the material being studied. The side-tube is connected either to a gas-burette, containing water or mercury, or to a graduated tube standing over water in a pneumatic trough, so that the volume

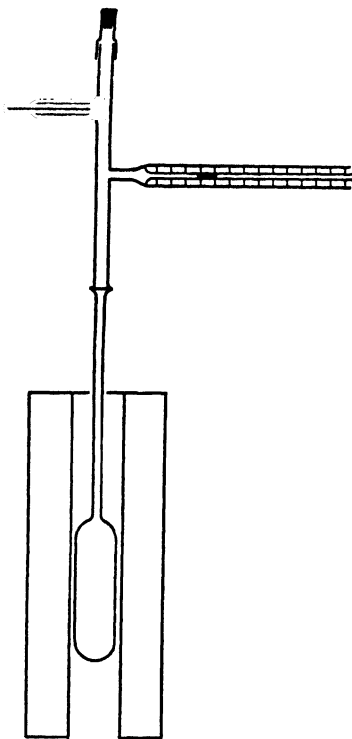


FIG. 52. Vapor densities at high temperatures (Nernst and von Wartenberg)

of air expelled from the main apparatus can be measured. When a steady state is reached, a weighed quantity of the experimental substance, contained in a Hofmann bottle, is introduced into the apparatus. The liquid in the bottle vaporizes and expels an equal volume of air which is measured at atmospheric pressure in the gas-burette or graduated tube already mentioned. A great advantage of the Victor Meyer method is that it is not necessary to know the temperature at which vaporization occurs. The vapor pushes out an equal volume of air at the experimental temperature, but this is cooled in the measuring apparatus and the volume recorded is that which the vapor would occupy if determined at the temperature of the gas-burette. The assumption made is that the vapor has the same coefficient of expansion as air, but since the method is approximate in any case, the error is not serious. As the weight of the material is known, and the volume of its vapor, at the atmospheric temperature and pressure, has been measured, the molecular weight can be calculated. Some authors suggest that the vapor pressure of the liquid, e.g., water, over which the expelled air is collected should be deducted from the barometric pressure, thus giving the pressure of the air; application of this correction implies that the air prior to expulsion was perfectly dry. To be strictly

accurate an allowance should be made for the *difference* in the pressures of water vapor in the air in the two portions of the apparatus, but since the results are only approximate it is probably best, unless the atmosphere is very dry, to ignore the correction entirely.

Modified Victor Meyer Methods.—For work at ordinary temperature the Victor Meyer apparatus is made of glass, but for higher temperatures, up to 1500° c., J. Mensching and V. Meyer (1886) employed porcelain, whereas the use of iridium (W. Nernst, 1903) permits operation up to 2000° c. Since the metal becomes permeable to gases at these temperatures, the apparatus must be covered with a suitable material; the best results were obtained by coating both inside and outside with magnesium oxychloride from a fused mixture

of the oxide and chloride (H. von Wartenberg, 1907). The necessary temperatures were obtained by an electric furnace, and since air could not be used inside the vessel, because of the possibility of chemical action, it was replaced by an inert gas, such as nitrogen or, better, argon. In the apparatus used by Nernst and von Wartenberg the capacity of the bulb was about 2 to 3 cc.; the experimental substance was weighed on a microbalance, and the volume of the expelled gas was determined by the movement of a drop of mercury in a graduated horizontal capillary attached to the side-tube (Fig. 52). Very important results concerning the molecular state of various substances at high temperatures, to which reference will be made shortly, were obtained in this way.

The Constant Volume Method.—The constant volume method for determining vapor densities was apparently first proposed by O. Bleier and L. Kohn (1899), and improved by J. S. Lumsden (1903); the latter used an apparatus similar to that of Victor Meyer, but with a much shorter tube, so that it consists essentially of a bulb of about 100 cc. capacity and a short vertical limb with a side-tube. To this tube is connected a mercury levelling device, similar to a gas-burette, to act as a manometer for measuring the pressure inside the apparatus. The bulb is surrounded by the vapor of a liquid boiling at a temperature sufficiently high to cause the experimental material to vaporize. In this method, as distinct from the Victor Meyer method, however, it is necessary to know the vaporization temperature. The side-tube of the apparatus is opened to the air by means of a stopcock, and when a steady state is reached the mercury is adjusted to a definite mark, representing the constant volume of the bulb. The stopcock is closed and a weighed quantity of the substance is introduced in the usual manner; it vaporizes and there is at once a tendency for the volume to increase, but this is prevented by raising the mercury levelling tube. When a steady condition is again reached and vaporization is complete, the mercury level is adjusted exactly to the mark and the increase of pressure is recorded; this represents the partial pressure of the vapor in the apparatus. If the volume of the bulb, up to the mark, and its temperature are known, the molecular weight of the vapor can be calculated. The Lumsden apparatus can be standardized by means of a substance of known molecular weight, so that the volume and temperature are not needed. Suppose a weight w_1 of substance of known molecular weight M_1 produces an increase of pressure P_1 with the temperature jacket at T ; then assuming ideal behavior $M_1 = w_1RT/P_1V$, where V is the constant volume of the bulb. For another material, molecular weight M_2 , a weight w_2 results in an increase of P_2 in the pressure, at the same temperature, in the same apparatus, that is at the same volume; hence $M_2 = w_2RT/P_2V$,

$$\therefore M_1/M_2 = w_1P_2/w_2P_1, \quad (241)$$

so that the value of M_2 may be obtained without a knowledge of the volume or temperature of the apparatus. The vapor pressure apparatus described by A. W. C. Menzies (see p. 629) may be readily adapted for the determination of vapor densities by the constant volume method.¹⁵

Other Vapor Density Methods.—Vapor density measurements have been employed for the study of dissociation of molecules in the gaseous state (p. 321), and constant volume methods have been particularly used in this connection. In the Victor Meyer apparatus the vapor tends to diffuse into the air above it, so that its partial pressure does not remain constant; since the extent of dissociation is often dependent on the pressure the results obtained may be of little

value. In the constant volume method, however, the pressure is the same throughout the vapor, and further it can be varied by using different quantities of material. Since much of the work on dissociation has been carried out at high temperatures and with substances liable to attack mercury, various forms of apparatus have been devised by M. Bodenstein (1909 *et seq.*), G. Preuner (1909 *et seq.*), and others, in which the apparatus is made entirely of glass or silica. The pressure is measured with either a glass or quartz spiral-manometer, a "sickle gauge," or a similar piece of apparatus; these generally involve some form of pointer which changes its position according to the pressure in the vessel. The external pressure is then adjusted until the pointer returns to its zero reading, when the applied pressure, measured with a mercury manometer, is the same as that inside the vessel.

As mentioned on p. 279, the rate of effusion of a gas at known pressure through a small orifice can be used for determining the density; alternatively, by keeping the temperature and pressure constant the densities of two gases may be compared.¹⁶

Results *: Elements.—Molecular weight determinations by vapor density methods have provided information of considerable interest and importance; as a general rule the results are in agreement with those expected from ordinary chemical considerations, and these require no further comment. It may be noted that in the vapor state metals, e.g., sodium, mercury, thallium, zinc, cadmium and lead, are almost entirely monatomic,[†] and so also are the inert gases; in these cases the molecular and atomic weights are identical. The molecular weights of the halogens are twice their respective atomic weights, provided the temperatures are not too high, so that they exist as diatomic molecules; at high temperatures dissociation occurs, as described below. Nitrogen and oxygen each have two atoms in the molecule, but the other elements in the fifth and sixth groups of the periodic classification form more complex molecules. At relatively low temperatures the following formulae have been found in the vapor state: P_4 , As_4 and Sb_4 in the fifth, and S_8 and probably Se_8 in the sixth group.

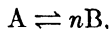
Association.—A number of compounds have molecular weights as vapors at atmospheric pressure which are higher than those to be expected from simple valence considerations. Such substances are said to be **associated** in the vapor state and to exhibit the phenomenon of **association**. For example, the formulae P_4O_{10} , P_4O_6 , As_4O_6 and Sb_4O_6 represent the oxides of the elements of the phosphorus group as vapors; the aluminum halides are Al_2Cl_6 and Al_2Br_6 , and ferric chloride is Fe_2Cl_6 . These values are for relatively low temperatures, for in each case dissociation into simpler molecules occurs on raising the temperature. The density of hydrogen fluoride vapor under normal conditions indicates the presence of $(HF)_n$ molecules, where n has a mean value of 6, but a simplification is evident at higher temperatures. The molecular weight of

* It should be understood, unless otherwise stated, that the results given are for atmospheric pressure.

† Small amounts of diatomic molecules, e.g., Na_2 , are often present in the vapor.

acetic acid vapor just above the boiling point at atmospheric pressure approaches twice the value to be expected from the chemical formula CH_3COOH , and formic acid shows a similar anomaly; the molecules are evidently $(\text{C}_2\text{H}_4\text{O}_2)_2$ and $(\text{CH}_2\text{O}_2)_2$, respectively (cf. p. 115).

Thermal Dissociation.—In many instances the vapor density, and hence the apparent molecular weight, of a substance decreases with increasing temperature; this is undoubtedly due to the splitting up of relatively complex into simpler molecules by heat, a phenomenon known as **thermal dissociation**. In physical chemistry the use of the term dissociation is to be distinguished from that of decomposition; the former always implies a process which can be reversed when the conditions are altered. In thermal dissociation, therefore, a molecule splits up into simpler parts on heating, but the latter recombine to form the original molecule on cooling. Vapor density measurements can be used to determine the extent of dissociation in the following manner. Consider the general reaction



that is, one molecule of A dissociates on heating into n molecules, which may be the same or different. If α is the degree of dissociation at any fixed temperature, that is, the fraction of the total number of molecules of A which are split up at equilibrium, then for each molecule of A present originally, $1 - \alpha$ molecules of A remain undissociated, while $n\alpha$ molecules of product have been formed, giving a total of $1 - \alpha + n\alpha$ molecules, for every one molecule expected theoretically. If w g. of a substance of molecular weight M occupied a volume V_i when no dissociation occurred, then assuming ideal behavior,

$$PV_i = \frac{w}{M} RT. \quad (242)$$

As a result of dissociation the number of moles is increased by a factor of $1 - \alpha + n\alpha$; the observed volume V_o occupied by the dissociated vapor is then given by

$$PV_o = \frac{w}{M} (1 - \alpha + n\alpha) RT. \quad (243)$$

From this equation it is possible to calculate the fraction of dissociation α at any temperature and pressure from a knowledge of the volume V_o occupied by a known weight of substance as vapor.

An alternative method of expressing the results is obtained if (243) is divided by (242), yielding

$$\frac{V_o}{V_i} = 1 - \alpha + n\alpha \quad (244)$$

at constant temperature and pressure. Since the density under these conditions is inversely proportional to the volume, and the apparent

molecular weight is proportional to the density, it follows that

$$\frac{\rho_t}{\rho_o} = \frac{M_t}{M_o} = 1 - \alpha + n\alpha, \quad (245)$$

where ρ_o and M_o are the theoretical density and molecular weight, and ρ_t and M_t are the observed values. If the dissociation is studied in a constant volume apparatus, the increase of pressure is inversely proportional to the density [cf. equation (241)], at constant temperature, so that

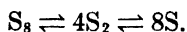
$$\frac{P_o}{P_t} = 1 - \alpha + n\alpha, \quad (246)$$

where P_o and P_t are the observed and theoretical pressure increases. Any of the foregoing expressions (243) to (246) may be used to calculate the degree of dissociation from vapor density, or related, measurements.

When a substance, e.g., acetic acid, is associated in the vapor, the degree of association can be determined from vapor density measurements by a method exactly similar to that given above.

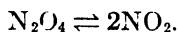
It is important to remember that the degree of dissociation is affected not only by temperature but also by pressure (see Chapter XI), and unless the pressure is explicitly stated it must be assumed to be equal to that of the atmosphere.

Dissociation of Elements.—At high temperatures the halogens are definitely dissociated into atoms at atmospheric pressure. At 1700° c. chlorine is about half dissociated, whereas iodine attains the same degree of dissociation at about 1200° c. The molecule of sulfur in the vapor, like that of the solid, is S_8 just above the boiling point, but the molecular weight decreases steadily from about 250 at 450° c., to 50 at 2070° c. (H. von Wartenberg, 1907). The latter value indicates that at this temperature the vapor contains some monatomic molecules, so that dissociation probably occurs in the stages



Some writers have suggested that S_6 and S_4 molecules are present at intermediate stages, but this is uncertain; the S_8 molecule has a ring structure (p. 377) and when it is broken it is probable that it will split mainly into S_2 molecules in the vapor, although S_6 and S_4 may exist in the liquid. The paramagnetism of sulfur vapor shows that S_2 has a structure similar to that of diatomic oxygen (p. 107). Selenium vapor on heating behaves similarly to sulfur; there is evidence of Se_6 or Se_8 molecules at low temperatures, and of monatomic molecules above 1000° c. Phosphorus, arsenic and antimony dissociate from P_4 , As_4 and Sb_4 to P_2 , As_2 and Sb_2 on heating; the presence of phosphorus and arsenic atoms in the vapor at temperatures above 1200° c. has been reported, but there appears to be some doubt about the accuracy of the claim. Antimony and bismuth are said to be almost entirely monatomic above 2000° c.

Dissociation of Compounds.—Aluminum chloride vapor, which at atmospheric pressure consists mainly of Al_2Cl_6 molecules at 200°C. , dissociates into single molecules on heating, and at 800°C. the vapor is almost entirely in this form. Similarly, aluminum bromide and ferric chloride molecules have the formulae AlBr_3 and FeCl_3 at high temperatures, for the double molecules dissociate on heating. The vapor density of arsenious oxide corresponds to As_4O_6 molecules at 600°C. and to As_2O_3 at 1800°C. ; in between these temperatures both forms are present in the vapor. An interesting case of dissociation is that of nitrogen tetroxide: at 22°C. , just above its boiling point, the vapor is pale reddish-brown in color and the molecular weight indicates that it consists mainly of N_2O_4 molecules. On heating, the vapor density decreases, while at the same time the color becomes darker until it is almost black; the molecular weight at 140°C. is 46, indicating complete dissociation into NO_2 molecules; thus,



Acetic acid vapor, which consists almost entirely of double molecules just above the normal boiling point, dissociates into simpler molecules on heating, dissociation being complete at 250°C. Decrease of pressure favors the dissociation, as is to be expected from the Le Chatelier principle (p. 831), and the results in Table 43, obtained at relatively low temperatures, are of interest in this connection (F. H. MacDougall, 1936).

TABLE 43. DEGREE OF DISSOCIATION OF $(\text{CH}_3\text{COOH})_2$ MOLECULES AT LOW PRESSURES

Pressure	25°	30°	35°	40°C.
5 mm.	0.165	0.209	0.252	0.294
10	0.112	0.149	0.182	0.218
15	0.083	0.119	0.150	0.185

In the examples of dissociation considered so far the molecules split up into simple molecules, or atoms, of the same kind; there are in addition important cases of dissociation into different molecules. It was discovered by J. B. A. Dumas (1836) that the density of ammonium chloride vapor was half the expected value, with the consequence that he, and others, were led to doubt the validity of Avogadro's hypothesis. The abnormal result was explained independently by H. Kopp, F. A. Kekulé and S. Cannizzaro (1857-58) as being due to dissociation on heating; thus,

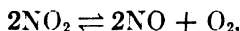


Since one molecule of ammonium chloride produces two molecules, (245) shows that when dissociation is complete the molecular weight will be half the theoretical value. Proof that ammonium chloride does actually produce ammonia and hydrogen chloride on heating was provided by L. Pebal (1862) and K. Than (1864), who achieved a partial separation of the gases by diffusion.

The vapor densities of phosphorus pentachloride and pentabromide are also less than the expected values, indicating dissociation into one

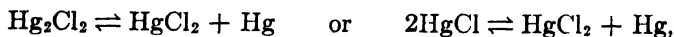
molecule of trihalide and one of halogen in each case; this view is supported by the color of the vapor, which resembles that of chlorine in the former case and that of bromine in the latter. The color increases in depth as the temperature is raised. The density measurements show that the degree of dissociation of the pentachloride is about 0.4 at 180° c. and is nearly complete at 300° c. at 1 atm. pressure.

It has been recorded above that N_2O_4 has completely dissociated into NO_2 at 140° c. and at the same time the gas has darkened in color. If the heating is continued the color becomes lighter, at first slowly and then more rapidly, and the change is accompanied by a further decrease of density; at 620° c. the gas is colorless and its molecular weight is about 30. These results show that above 140° c. the deeply colored nitrogen dioxide dissociates into colorless nitric oxide and oxygen, i.e.,



and the process is complete at about 620° c. The density and color changes are reversed on cooling.

The vapor density of mercurous chloride suggests the formula $HgCl$, but it has been proved that the vapor contains free mercury and mercuric chloride. It is clear, therefore, that dissociation occurs on heating; this may be either



for in each case the density would be about half the value required for Hg_2Cl_2 . In view of other evidence that the mercurous ion is Hg_2^{++} , and not Hg^+ , it is probable that the first alternative is the correct one.

In addition to the instances already mentioned, the dissociation of the following substances, among others, has been studied by density methods, or by measurements of increase of pressure at constant volume which may be regarded as the same in principle: antimony pentachloride, sulfuryl chloride, carbonyl chloride and bromide, iodine monobromide and carbon dioxide.

It has been stated in the literature that perfectly dry ammonium chloride and mercurous chloride do not dissociate on heating, but the matter is in doubt; the subject is considered more fully on p. 1129.

In conclusion, it is important to point out that whereas a molecular weight lower than the theoretical implies dissociation, the process of dissociation is not always accompanied by an alteration of density. This only occurs when there is a change in the number of molecules as a result of dissociation. When hydrogen iodide dissociates, for example, there is no such change, and the density measurements provide no information on the extent of dissociation.

HEAT CAPACITIES OF GASES

Molar Heat Capacities.—The difference in the molar heat capacities at constant pressure and constant volume of an ideal gas (p. 195) is given

by

$$C_P - C_V = R. \quad (247)$$

The same result may be derived by utilizing the fact that for an ideal gas the difference in the heat required to raise the temperature at constant pressure and at constant volume is equal to the work done in expansion (p. 194). If V_1 is the volume of 1 mole of gas at T_1 and V_2 that at T_2 , the pressure P being constant, then

$$\begin{aligned} C_P(T_2 - T_1) - C_V(T_2 - T_1) &= P(V_2 - V_1); \\ \therefore (C_P - C_V)(T_2 - T_1) &= PV_2 - PV_1. \end{aligned} \quad (248)$$

Since the gas is ideal,

$$\begin{aligned} PV_1 &= RT_1 \quad \text{and} \quad PV_2 = RT_2; \\ \therefore PV_2 - PV_1 &= R(T_2 - T_1), \end{aligned} \quad (249)$$

and combination of (248) and (249) gives

$$C_P - C_V = R.$$

The difference between the molar heat capacities at constant pressure and constant volume for an ideal gas should be 1.987 cal. deg.⁻¹.

According to the kinetic theory, the kinetic (translational) energy of 1 mole of an ideal gas is $\frac{3}{2}RT$ [equation (16)], and if this is *the only kind of energy* possessed by such a gas, it is possible to write

$$E_K = E = \frac{3}{2}RT. \quad (250)$$

The energy content of an ideal gas is dependent only on its temperature (p. 191), and so differentiation of (250) with respect to temperature at constant volume gives

$$C_V = \left(\frac{\partial E}{\partial T} \right)_V = \frac{3}{2}R. \quad (251)$$

This result can also be deduced by equating the heat absorbed at constant volume, i.e., $C_V(T_2 - T_1)$, to the increase of kinetic energy of the molecules, $\frac{3}{2}R(T_2 - T_1)$. The same assumption is involved as before, namely, that the translational energy is the only kind of energy, and all the heat supplied at constant volume is utilized in increasing the kinetic energy, i.e., the speed, of the molecules.

Since C_V is equal to $\frac{3}{2}R$, and $C_P - C_V$ is equal to R , it follows that C_P is $\frac{5}{2}R$; the ratio of the molar, or specific, heats of an ideal gas, represented by the symbol γ , should thus be

$$\gamma = \frac{C_P}{C_V} = \frac{5}{3} = 1.666. \quad (252)$$

It is of interest to see how far the observations made with real gases comply with the requirements for an ideal gas. The experimental methods fall into two categories, viz., those involving separate measurements

of heat capacities at constant pressure and constant volume, and those in which the ratio only is determined.

Constant Volume Methods.—Accurate determinations of specific heats at constant volume were made by J. Joly (1890) with the aid of his steam calorimeter. The gas under examination is placed in a copper sphere which is counterpoised by a similar, but evacuated, one. Steam is allowed to condense on the two spheres, and from the difference in the amounts of water formed, and the known latent heat of vaporization, the heat capacity of the gas at constant volume can be calculated. In the method of R. Bunsen (1850) a known quantity of the gas under examination is mixed with a definite amount of detonating gas, i.e., a mixture of two volumes of hydrogen to one of oxygen, and placed in a steel bomb. The gases are exploded by a spark, and the maximum pressure attained is measured; from this the temperature can be calculated on the assumption that the gas laws are applicable, allowance being made for the hydrogen and oxygen removed by reaction. From a knowledge of the heat evolved when hydrogen and oxygen combine, the heat capacity of the experimental gas can be determined, provided that of the water vapor formed is known. An important improvement in the explosion method was made by M. Pier (1908), who used a photographic device for measuring the rapid increase of pressure following the explosion. The Bunsen method is, of course, only applicable to gases which do not react chemically with hydrogen or oxygen under the experimental conditions.

For measurements at very low temperatures A. Eucken (1912) has employed a method in which the heat capacity of a thin-walled steel vessel is determined, first when evacuated and then when filled with gas under pressure. A constantan wire is wound round the vessel so that a known amount of heat can be supplied electrically; the resulting increase of temperature is measured by a resistance thermometer. In order to reduce heat losses the apparatus is suspended in an evacuated space. The procedure described can only be used for low temperatures, e.g., below -100°C ., when the heat capacity of the vessel itself is small in comparison with that of the gas contained therein.

Constant Pressure Methods.—In H. V. Regnault's method (1862) a known quantity of gas flows, under a measured constant pressure, through a copper spiral in a warm constant temperature bath, and then through another spiral immersed in a water calorimeter of known thermal capacity. From the increase in temperature of the latter the specific heat of the gas can be evaluated. In a modified procedure, the gas is passed through a bath kept at a convenient low temperature, and when passed into the calorimeter cooling is avoided by electrical heating in amount just sufficient to maintain a constant temperature. For work at high temperatures, up to 1400°C ., L. Holborn and F. Henning (1907) passed the gas first through an oven to receive a preliminary heating and then through a jacketed platinum tube where it is heated electrically to a definite temperature measured by a thermocouple. The gas then passes through three silver cylinders, in series, packed with silver filings to provide good thermal contact; the cylinders are contained in a calorimeter in which the liquid is paraffin oil. The heat capacity of the calorimeter and its contents are determined by the electrical method (p. 208).

The continuous flow method, used by H. L. Callender and H. T. Barnes in their classical work on the specific heat of water, has been adapted by W. F. G.

Swann (1909) for measurements with gases at constant pressure; it has also been used in the accurate work of K. Scheel and W. Heuse (1912). The gas, at a known temperature measured on a resistance thermometer, flows upward at a known rate over an electrical heater supplying heat at a definite rate, and the temperature is then measured by another resistance thermometer. Before reaching the heater the gas is circulated through an annular space surrounding it, so that heat escaping from the heated gas is not lost but utilized by the on-coming gas. From a knowledge of the rates of flow of gas and of supply of heat, and the initial and final gas temperatures, the specific heat of the gas may be determined at constant pressure.

In another form of the flow method (P. M. S. Blackett, *et al.*, 1930) the gas is passed at a slow rate through a narrow metal tube, the ends of which, initially at the same temperature, are connected to thermojunctions while the middle portion is heated by an electric current. A difference of temperature is set up between the ends of the tube, depending on its thermal conductivity and dimensions, and also on the specific heat and conductivity of the gas. The apparatus is standardized by means of a gas of known specific heat.

Ratio of Specific Heats.—In an adiabatic expansion from initial pressure P_1 and volume V_1 to pressure P_2 and volume V_2 , the relationship $P_1V_1^\gamma = P_2V_2^\gamma$ applies (p. 199), and this equation, in conjunction with Boyle's law, was used by F. Clément and C. B. Desormes (1812) to determine the ratio of the specific heats. The gas at a pressure greater than atmospheric is placed in a 30–50 liter vessel provided with a stopcock and a water or oil manometer. The pressure P_1 is measured, the stopcock is suddenly opened, to allow the pressure to fall to that of the atmosphere, and then closed; during the expansion, which is virtually adiabatic, the gas is cooled and as it warms up to its original temperature the pressure rises to P_2 . If V_1 is the volume of 1 mole of gas at pressure P_1 , and V the corresponding volume after adiabatic expansion when the pressure is P , then for an ideal gas

$$P_1V_1^\gamma = PV^\gamma. \quad (253)$$

Since the initial and final temperatures, when the pressure has become P_2 , are the same, and the corresponding volumes of 1 mole of gas are V_1 and V respectively, it follows, by Boyle's law, that

$$P_1V_1 = P_2V. \quad (254)$$

Eliminating V_1 and V from (253) and (254) there results

$$\begin{aligned} (P_1/P_2)^\gamma &= P_1/P; \\ \therefore \gamma &= \frac{\log P_1 - \log P}{\log P_1 - \log P_2}. \end{aligned} \quad (255)$$

From the three pressure measurements, therefore, the ratio of the specific heats can be calculated.

An improvement on the method was made by O. Lummer and E. Pringsheim (1898) and by J. R. Partington (1913, 1921). A large copper globe, over

100 liters capacity, is used, and the stopcock is allowed to remain open after adiabatic expansion. The temperatures before (T_1) and after expansion (T_2) are measured by a sensitive resistance thermometer. Using the same symbols as previously, it follows that $P_1 V_1^\gamma = P V^\gamma$, as before; further

$$P_1 V_1 = R T_1 \quad \text{and} \quad P V = R T_2,$$

and eliminating V_1 and V , it is seen that

$$(T_1/T_2)^\gamma = (P_1/P)^{\gamma-1}; \quad (256)$$

$$\therefore \gamma = \frac{\log P_1 - \log P}{\log (P_1/P) - \log (T_1/T_2)}. \quad (257)$$

For accurate results correction must be made for departure of the gas from ideal behavior, and for this purpose the Berthelot equation has been found convenient.

Measurement of the velocity of sound has been utilized for the determination of the ratio of specific heats. On account of the rapidity of sound vibrations, the successive compressions and expansions are adiabatic, and it can be shown that

$$\gamma = c^2 \left(\frac{\partial \rho}{\partial P} \right)_T, \quad (258)$$

where c is the velocity of sound and ρ is the density of the gas. For an ideal gas, since ρ is proportional to P at constant temperature, (258) becomes

$$\gamma = c^2 \rho / P = c^2 M / RT, \quad (259)$$

where M is the molecular weight of the gas. For a real gas it is necessary to express the density in terms of the pressure before differentiation can be carried out, and the Berthelot equation may be used. Since $1 + x$ is practically equal to $1/(1 - x)$ when x is small, it follows from (169) that

$$\rho = \frac{M}{V} \approx \frac{MP}{RT} \left[1 - \frac{9}{128} \cdot \frac{PT_c}{P_c T} \left(1 - 6 \frac{T_c^2}{T^2} \right) \right]; \quad (260)$$

$$\therefore \gamma = c^2 \left(\frac{\partial \rho}{\partial P} \right)_T = \frac{c^2 M}{RT} \left[1 - \frac{9}{64} \cdot \frac{PT_c}{P_c T} \left(1 - 6 \frac{T_c^2}{T^2} \right) \right] \quad (261)$$

The velocity of sound can be measured directly by means of a long coiled tube, e.g., 15 meters long and 2.5 cm. diameter; one end is closed by a steel diaphragm, so that on being struck by a hammer an impulse is sent along the tube. Small platinum foil diaphragms are placed over holes near each end, and the time the sound wave passes them is recorded electrically on a chronograph. Knowing the length of the tube, the velocity of sound, and hence the ratio of the specific heats, can be calculated. The method has been used for measurements up to 1000°C . (H. B. Dixon, 1921). A more common procedure is to use the principle proposed by A. Kundt (1868) in which the wave length of sound of known frequency is measured in the gas; if ν is the frequency and λ the wave length, then c is equal to $\lambda \nu$. In the original method a thin layer of lycopodium or other fine powder is spread along the tube, and the gas in it is set vibrating by stroking, with damp chamois leather, a glass rod terminating in a disc which nearly fills the cross section of the tube. The powder forms itself in

small heaps at the nodes of the sound waves, so that the distance between two heaps is half a wave length. In order to avoid direct measurement of the frequency of the vibrations the apparatus is calibrated with air, in which the velocity of sound is accurately known. U. Behn and H. Geiger (1907) modified the method by enclosing the gas under investigation in a sealed tube containing the fine powder; the tube is clamped in the middle and vibrations are set up by rubbing it in the usual manner. The frequency of vibration is varied by attaching metal discs to the ends of the tube until the length corresponds to an exact number of half wave lengths. One end of the experimental tube projects into a slightly wider tube containing air, so that vibrations of the same frequency are set up in the air for calibration purposes.

In the procedure adopted by J. R. Partington and W. G. Shilling (1923) for measurements up to 1000°C . the use of powder is avoided. One end of a tube, about 1 meter long, is closed by a telephone receiver which is actuated by an oscillator and emits a note of constant known frequency; at a distance of slightly less than half a wave length a side tube is inserted and to this is attached a listening device. The other end of the long tube, containing the gas, is closed with a gas-tight movable piston. This is gradually pushed in and a series of positions found at which the sound heard in the listening tube indicates a sharp maximum; the distance between any two successive positions gives the half wave length of the sound under the particular conditions.¹⁷

The velocity of sound measured in tubes differs from that obtained for a free space, and an appropriate correction must always be applied to the former. If c' is the velocity determined in a tube, and c the true velocity, then

$$c' = c(1 - ab),$$

where a is a constant for the tube, and b is a factor, varying with temperature, which depends on the thermal conductivity, viscosity, specific heat and density of the gas (G. R. Kirchhoff, 1868).

Experimental Results: Difference of Molar Heat Capacities.—The experimental data will now be considered from several aspects; the results in Table 44 give the values for the molar heats of some gases at ordinary temperatures, and the differences between them. Although the differ-

TABLE 44. MOLAR HEAT CAPACITIES OF GASES AND THEIR DIFFERENCES

Gas	C_P	C_V	$C_P - C_V$
Hydrogen	6.871 cal. deg. ⁻¹	4.884 cal. deg. ⁻¹	1.987 cal. deg. ⁻¹
Helium	5.004	3.017	1.987
Oxygen	6.989	4.994	1.995
Carbon monoxide	7.014	5.017	1.997
Carbon dioxide	8.89	6.86	2.03
Ammonia	8.77	6.70	2.07
Ethylene	10.23	8.19	2.04

ences between the molar heats are close to the theoretical value for an ideal gas, nevertheless appreciable deviations occur; the discrepancies are more marked at low than at high temperatures and are particularly evident for easily liquefiable gases, as the results in Table 44 indicate. Since these are the conditions under which gases depart from ideal be-

havior, it is probable that this is the cause of the failure of the measured difference between the molar heat capacities of a gas to equal the theoretical value. The deduction on p. 323 assumes that the internal energy of the gas is independent of its volume, but on account of molecular attractions this is not true for a real gas.

If the van der Waals equation is written in the form of (146), p. 291, to obtain $(\partial P/\partial T)_V$,

$$P = \frac{RT}{V-b} - \frac{a}{V^2},$$

application of the thermodynamic equation (p. 236), for 1 mole,

$$C_P - C_V = T \left(\frac{\partial P}{\partial T} \right)_V \left(\frac{\partial V}{\partial T} \right)_P, \quad (262)$$

which applies to any substance, gives

$$C_P - C_V = \left(P + \frac{a}{V^2} \right) \left(\frac{V-b}{T} + \frac{2a}{RT^2} \right), \quad (263)$$

utilizing the value of $(\partial V/\partial T)_P$ derived on p. 292. Multiplying out (263) and neglecting the term involving a^2 , the result

$$C_P - C_V = R + \frac{2aP}{RT^2} \quad (264)$$

is obtained. It is immediately clear that for a real gas $C_P - C_V$ must be greater than the theoretical value, and the discrepancy should be greater the higher the pressure and the lower the temperature; further, the differences are most marked when a is large, that is, for easily liquefiable gases. The treatment given here may also be applied to the Berthelot equation (p. 295). Starting with (168), and replacing a , b and R by their empirical equivalents in terms of the critical temperature and pressure, it is found that

$$C_P - C_V = R \left(1 + \frac{27}{16} \cdot \frac{PT_c^3}{P_c T^3} \right). \quad (265)$$

By applying a correction of the type required by (264), to allow for departure from ideal behavior, K. Scheel and W. Heuse (1912-13) found the differences between the molar heat capacities of gases, even at low temperatures, to be in agreement with theoretical requirements. The discrepancy between the experimental and theoretical values of $C_P - C_V$ for real gases can thus be attributed to their deviation from ideality.

A more serious discrepancy between real and ideal behavior is the fact that, except for helium, the results in Table 44 indicate larger molecular heats than those calculated on p. 323, viz., $\frac{3}{2}R = 2.98$ cal. for C_V , and $\frac{5}{2}R = 4.97$ cal. for C_P . The greater the number of atoms in the molecule, in general, the higher are the molar heat capacities. This fact suggests that the high results are to be attributed to the neglect of part of the energy in the calculations given above. The more complex the molecule the larger is likely to be the internal energy of rotation and vibration

The energy E of a molecule may be divided into two parts, viz., the translational energy $E_{tr.}$, due to the movement (translation) of the molecule as a whole, and the internal energy $E_{int.}$, due to the vibrational and rotational motions within the molecule itself; hence, it is possible to write

$$E = E_{tr.} + E_{int.} \quad (266)$$

The translational (kinetic) energy $E_{tr.}$ is identical with the quantity which has previously been represented by the symbol E_K ; this is equal to $\frac{3}{2}RT$ per mole, so that

$$E = \frac{3}{2}RT + E_{int.};$$

$$\therefore C_V = \left(\frac{\partial E}{\partial T} \right)_V = \frac{3}{2}R + \left(\frac{\partial E_{int.}}{\partial T} \right)_V. \quad (267)$$

The quantity $(\partial E_{int.}/\partial T)_V$ may be called the internal molar heat capacity, $C_{int.}$, so that

$$C_V = \frac{3}{2}R + C_{int.}, \quad (268)$$

and taking $C_P - C_V$ to be equal to R , which is at least approximately true, (268) gives

$$C_P = \frac{5}{2}R + C_{int.}. \quad (269)$$

It is thus the inclusion of internal energy, which increases with the complexity of the molecule, that accounts for the observed heat capacities being greater than the translational values, i.e., $\frac{3}{2}R$ and $\frac{5}{2}R$, except for helium and the other inert (Group 0) gases.

Equipartition of Energy.—The classical method of calculating $C_{int.}$ is based on the principle of the equipartition of energy. It was shown on p. 249 that the square of the component of molecular velocity in one particular direction is equal to one-third of the square of the actual velocity, and hence it is reasonable to assume that the translational energy can be regarded as made up of three equal parts in directions parallel to three specified axes at right angles. The total translational energy per mole is $\frac{3}{2}RT$, and so the value in each direction will be $\frac{1}{2}RT$ per mole or $\frac{1}{2}kT$ per single molecule, where k is the Boltzmann constant. This result is generalized in the equipartition principle which may be stated in the following form: *the total energy of a molecule is divided equally among the different degrees of freedom*, the amount being $\frac{1}{2}kT$ for each. In the present connection the expression **degrees of freedom** refers to the number of independent "square terms," i.e., terms involving the square of a coordinate or momentum (velocity), required to express the total energy of the molecule. The translational energy of a molecule can be expressed as the sum of three square terms, viz., $\frac{1}{2}mu^2$, $\frac{1}{2}mv^2$ and $\frac{1}{2}mw^2$, where u , v and w are the velocity components, so that if each square term contributes $\frac{1}{2}kT$ to the energy of the molecule, the total translational energy should be $\frac{3}{2}kT$ per molecule, as given above. Further, rotational and vibrational energy must be considered. The energy of rotation is kinetic in nature and is dependent only on the square of the angular momentum; hence

each type of rotation which varies with temperature contributes one square term, and consequently $\frac{1}{2}kT$ toward the energy. On the other hand, vibrational energy is partly kinetic and partly potential; each is expressed by a square term, and so each mode of (harmonic) vibration should contribute kT to the energy of the molecule.

The molecules of a monatomic gas appear to possess energy of translation only. Rotational energy is probably present, but this is of a type which does not vary with temperature and so can be ignored. Further, a single atom has no vibrational energy, so that the total number of square terms is three, and the total energy is $\frac{3}{2}kT$ per molecule or $\frac{3}{2}RT$ per mole. The molar heat capacities of a monatomic gas, therefore, do not involve any internal contributions, and the values should correspond to those for an ideal gas. There is good reason for believing that the inert gases of the atmosphere are monatomic, and experiments have shown that their molar heat capacities are very close to 5 cal. and 3 cal. deg.⁻¹ at constant pressure and constant volume respectively; the ratio is 1.666, which is in accordance with the requirements for an ideal gas. It is generally accepted, therefore, that a gas for which γ is 1.66, independent of temperature,* is actually monatomic, and measurements of specific heats have led to the conclusion that the vapors of mercury and of the alkali metals are largely monatomic. Spectroscopic and other measurements show, however, that alkali metal vapors contain a small proportion of diatomic molecules.

When there are two atoms in the molecule, connected somewhat in the manner of a dumbbell, rotation about two axes at right angles is possible. Consider three mutually perpendicular axes, XX , YY and ZZ in Fig. 53; the first two

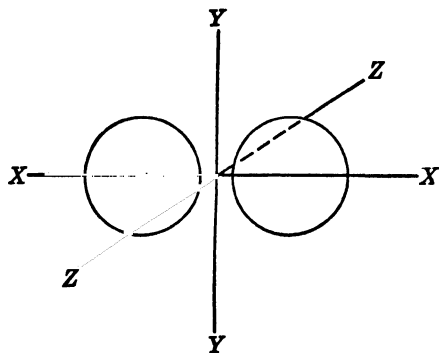


FIG. 53. Rotation of a diatomic molecule

are in the plane of the paper and the other at right angles. Rotation about the axis XX , which passes symmetrically through the atoms, is analogous to that of a monatomic molecule, and so does not influence the heat capacity. The rotations of the molecule about YY and ZZ each contribute one square term, and hence according to the equipartition principle the rotational

energy should be kT per molecule. With a diatomic molecule vibration is also possible, since the two atoms may oscillate with respect to each

* At sufficiently low temperatures, when all the molecules are in their lowest rotational and vibrational energy states, the value of γ should be 1.66 for *any* gas; this has been observed for hydrogen and deuterium. However, in these cases γ decreases as the temperature is raised, as other forms of energy, in addition to translational, contribute to the heat capacity.

other along the XX axis; if such vibration should occur the contribution to the total energy will be a further kT per molecule, since two square terms are concerned.

The number of energy square terms for a polyatomic molecule depends on its configuration. There are always three for translation, and except for a linear molecule which has only two (cf. Fig. 53), there are three rotational square terms. In general, a molecule containing n atoms has $3n - 6$ normal modes of vibration, although for a linear molecule this is increased to $3n - 5$;^{*} of these $n - 1$ are "stretching" or "valence bond" vibrations, and the remainder are "deformation" or "bending" vibrations. Each mode of vibration should contribute two square terms to the total energy.

The conclusions reached from the principle of equipartition of energy are best considered in the light of the experimental facts. A number of data have therefore been collected in Table 45, in which the second column gives the ratio of the specific heats at ordinary temperatures, and the others contain the molar heat capacities at constant volume for a number of temperatures up to 2000° c.

TABLE 45. MOLAR HEAT CAPACITIES AT CONSTANT VOLUME IN CAL. DEG.⁻¹

Gas	γ	0°	100°	200°	500°	1200°	2000° c.
A	1.666	2.98	2.98	2.98	2.98	3.00	3.00
H ₂	1.41	4.87	4.93	5.05	5.16	5.67	6.28
N ₂ , O ₂ , CO	1.40	4.99	5.05	5.15	5.26	5.75	6.3
HCl	1.39	5.00	5.09	5.27	5.46	6.13	6.9
Cl ₂	1.36	5.95	6.3	6.7	6.9	7.1	7.2
H ₂ O	1.31	5.93	6.00	6.60	7.00	8.4	11.0
CO ₂	1.31	6.68	7.69	9.04	9.75	10.6	11.1
NH ₃	1.32	6.62	7.05	8.3	9.5	11.4	—

The results for the monatomic inert gases are in agreement with theory, and so need not be discussed further, but the figures for diatomic gases are worthy of consideration. The molar heats at constant volume of hydrogen, nitrogen, carbon monoxide, hydrogen chloride, and also of nitric oxide and hydrogen bromide and iodide, are all approximately 5 cal. deg.⁻¹ at ordinary temperatures. Subtracting 3 cal. for the translational contribution leaves 2 cal. deg.⁻¹ for the internal molar heat; this means the internal energy, i.e., of rotation and vibration, is RT per mole or kT per molecule. There are thus two square terms of internal energy, implying either two rotations or one vibration; assuming the former to be the case, these diatomic molecules would be regarded as rigid rotators. The molar heat of chlorine at constant volume is about 6 cal. deg.⁻¹, and so the internal energy must be $\frac{3}{2}RT$ per mole, since the internal molar heat is $\frac{3}{2}R$, i.e., approximately 3 cal. deg.⁻¹. The internal energy is consequently

* When free rotation is possible within the molecule the rotational square terms are increased by one, and the modes of vibration are decreased by one for each type of rotation. If the rotation is restricted, as in ethane, special treatment is necessary (p. 877).

distributed among three square terms; this is difficult to understand, unless there is rotation about only one axis, in addition to vibration. Since the two types of rotation are really identical (see Fig. 53), it is improbable that energy from one only will be concerned in the specific heat. A possible explanation is one which was proposed to account for the fact that molar heat capacities do not increase in steps of $\frac{1}{2}R$ as the temperature is raised. It was assumed that only in a proportion of the molecules do all the available square terms contribute to the energy, and the number changes with temperature. For example, if in chlorine at ordinary temperatures half the molecules rotated and vibrated, giving a total of seven square terms, including those of translation, whereas the other half rotated but did not vibrate, giving five square terms, the heat capacity of the mixture would be 6 cal. deg.⁻¹. Increase of temperature would result in an increased proportion of the vibrating molecules and so the value should rise to a maximum of 7 cal. deg.⁻¹; it is of interest to note, however, that this value is exceeded.

Quantum Theory of Heat Capacity.—A more satisfactory treatment of the heat capacities of gases, which shows the equipartition principle to be an approximation applicable in certain limiting cases, is based on an application of statistical methods in conjunction with quantum theory.

The treatment to be described here is applicable to all forms of energy, which, according to quantum theory, are not infinitely variable, but can have only certain definite values. The method used is a development of that indicated on p. 96 in connection with rotational energy. The number of molecules in a quantum state, in which the energy is ϵ_i in excess of the zero-point amount, is, according to the classical distribution law (p. 273), equal to $p_i N_0 e^{-\epsilon_i/kT}$, where p_i is the statistical weight, or multiplicity, of the given state (p. 95), and N_0 is the number of molecules in the lowest energy state. When 1 mole of gas is considered the total number of molecules is equal to the Avogadro number N , and this must be equal to the sum of the molecules in the various energy states, characterized by 0, 1, 2, etc.; thus (cf. p. 32),

$$N = p_0 N_0 + p_1 N_0 e^{-\epsilon_1/kT} + p_2 N_0 e^{-\epsilon_2/kT} + \dots \quad (270)$$

$$= N_0 \sum_{i=0}^{\infty} p_i e^{-\epsilon_i/kT}. \quad (271)$$

The energy, in excess of the zero-point value, of the molecules in any state is obtained by multiplying the energy of each molecule in a given state by the number in that state, so that the total energy per mole, referred to the zero-point energy, is

$$E = 0 p_0 N_0 + \epsilon_1 p_1 N_0 e^{-\epsilon_1/kT} + \epsilon_2 p_2 N_0 e^{-\epsilon_2/kT} + \dots \quad (272)$$

$$= N_0 \sum_{i=0}^{\infty} \epsilon_i p_i e^{-\epsilon_i/kT}. \quad (273)$$

Eliminating N_0 from (271) and (273), and omitting subscripts and summation

limits in order to simplify the representation,

$$E = \frac{N \sum \epsilon p e^{-\epsilon/kT}}{\sum p e^{-\epsilon/kT}} = RT^2 \frac{d \ln Q}{dT}, \quad (274)$$

where Q , known as the partition function of the molecule, is defined by

$$Q = \sum p_i e^{-\epsilon_i/kT}, \quad (275)$$

the sum being taken over all the energy states of the molecule.

The molar heat capacity may be obtained by differentiating the expression (274), for the total energy, with respect to temperature at constant volume; thus,

$$C_V = \left(\frac{\partial E}{\partial T} \right)_V = \frac{\partial}{\partial T} \left[RT^2 \left(\frac{d \ln Q}{dT} \right) \right]_V \quad (276)$$

$$= \frac{R}{T^2} \left[\frac{Q''}{Q} - \left(\frac{Q'}{Q} \right)^2 \right]_V, \quad (276a)$$

where Q' represents $dQ/d(1/T)$ and Q'' is $d^2Q/d(1/T)^2$.

The energy values ϵ to be used in the calculation of the partition function are, strictly speaking, the total energies in each quantum state of the molecule. If there is no interaction between the various forms of energy, so that the total energy in each state is the sum of the translational and internal, i.e., rotational, vibrational and electronic, energies, the complete partition function Q is then the product of the contributions from the translational, i.e., $Q_{tr.}$, and internal, i.e., $Q_{int.}$, degrees of freedom; thus,

$$Q = Q_{tr.} Q_{int.}$$

Similarly, $Q_{int.}$ may be taken as the product of $Q_{rot.}$, $Q_{vib.}$, and, if necessary, $Q_{el.}$, the rotational, vibrational and electronic contributions. All of these partition functions are defined by an expression similar to (275), but the energy values are restricted to one particular type in each case. Since E and C_V , and other thermodynamic properties, are determined by $\ln Q$ [cf. equations (274) and (276)], it is evident that they may be taken as equal to the sum of the contributions from the various forms of energy, obtained by applying (274) and (276) to the individual partition functions. For example, the translational partition function of a gas molecule is $(2\pi mkT)^{3/2} V/h^3$, where m is the mass of the molecule, h is Planck's constant, and V is the volume of the gas (p. 873). Inserting this expression for $Q_{tr.}$ into (274) and (276), it is found that the molar translational energy is $\frac{3}{2}RT$, while the heat capacity is $\frac{3}{2}R$. In this case the results are identical with the classical values, because the quanta of translational energy are so small that the energy varies virtually in a continuous manner.

In order to use (276) or (276a) to calculate the internal heat capacity of a gas, the essential point is the determination of the partition function $Q_{int.}$ for internal degrees of freedom at a number of temperatures. To do this it is necessary to evaluate the $p e^{-\epsilon/kT}$ terms and to sum them over *all possible internal energy states* of the molecule. The energy of each state can generally be obtained from spectroscopic data, since by the quantum theory $\epsilon = h\nu = hc\bar{\nu}$, where $\bar{\nu}$ is the wave number corresponding to a definite energy transition; care is necessary, however, in assessing the correct statistical weight to be applied to

any given energy state. For atoms and monatomic molecules the total multiplicity is the product of the values due to angular momentum of the nucleus, i.e., $2i + 1$, where i is the nuclear spin quantum number (p. 93), and to the resultant angular momentum of the electrons, i.e., $2j + 1$, where j is the so-called "inner" quantum number. In deriving the partition function every possible electronic level should, theoretically, be taken into consideration, but in actual practice, except for high temperatures, only the lowest energy levels are appreciably occupied and the others may be ignored. When the ground term of an atom is a multiplet, however, as is the case with the alkali metals and other elements (pp. 41-66), then the contribution of each of the sublevels must be included in the partition function.

Diatomic Gases.—From the standpoint of heat capacities the main interest, for present purposes, is attached to the study of diatomic molecules. Many of these have singlet electronic levels, although oxygen and nitric oxide have multiplets for which allowance must be made. At high temperatures, particularly for oxygen, it may also be necessary to include higher, i.e., excited, electronic states. Each vibrational level has a statistical weight of unity, and since the quanta of vibrational energy are relatively large, compared with those for rotation, it is rarely necessary to include a large number of terms in the partition function. For example, with hydrogen the first vibrational level only need be considered up to about 900°K. ; that is to say, the vibrational quantum is so large that there are only a negligible number of molecules in any level above the lowest. At 2000°K. , however, the contributions of three further levels must be included; the factor $e^{-\epsilon/kT}$, with the appropriate value of ϵ , shows that there are appreciable numbers of molecules in these levels, but extremely few in the fifth and higher levels. With nitric oxide and oxygen, on the other hand, the vibrational quanta are somewhat smaller, and five levels are occupied to an appreciable extent at 900°K. , and ten at 2000°K. , although at ordinary temperatures the lowest level only makes any appreciable contribution to the partition function. The vibrational quantum of molecular chlorine is very small, and so even at room temperatures a number of vibrational levels are occupied.

The quantum treatment permits a direct interpretation of the apparent anomalies of Table 45. At ordinary temperatures, the molecules of hydrogen, oxygen, nitric oxide, hydrogen chloride, and of other substances in which the atoms are firmly bound, are almost all in the lowest vibrational level; they consequently have no vibrational energy relative to the zero-point value and so behave as if they were rigid. As the temperature is raised, however, the molecules commence to occupy higher vibrational levels in considerable numbers, and these contribute to the heat capacity. With chlorine, on the other hand, a large number of molecules have appreciable vibrational energy in excess of the zero-point value even at normal temperatures, and hence its internal energy and heat capacity are higher than for the gases previously mentioned.

The question of the rotational statistical weights of diatomic molecules has been considered on p. 96 with special reference to hydrogen, and the rotational partition function for this substance can be written

$$Q_{\text{rot.}} = \sum_{J=0,2,4,6,\dots} (2J+1)e^{-J^2/kT} + 3 \sum_{J=1,3,5,7,\dots} (2J+1)e^{-J^2/kT} \quad (277)$$

where the first term refers to the para-form, with a nuclear spin statistical weight of unity, and the second to the ortho-form with a statistical weight of three. With certain molecules, e.g., oxygen and nitric oxide, various other multiplicities arise, and although the calculation of the partition function is complicated, it can nevertheless be carried out if the necessary spectroscopic data are available.

Heat Capacity of Hydrogen.—The determination of the complete partition function is a very tedious matter, because the rotational energy must be evaluated for each vibrational level, and the results added together, and the same calculation must be made for every electronic level. The labor can be reduced, however, by various mathematical devices. For hydrogen the partition function, and hence the internal energy and the heat capacity, can be determined without great difficulty. As mentioned above, vibrational energy need not be considered at temperatures below about 300° K., and under these conditions not more than seven rotational levels are occupied. If the rotational partition function is evaluated by (277), that is, for equilibrium-hydrogen in which the ortho-para ratio is assumed continuously to adjust itself to the equilibrium value, the results differ from those determined experimentally at low temperatures. This discrepancy presented a serious problem, until D. M. Dennison (1927) showed the explanation to lie in the fact that nuclear spins are not easily reversed; hence the hydrogen which is being studied does not contain the ortho- and para-forms in the equilibrium amounts, but in the ratio present in normal hydrogen, i.e., three parts of ortho- to one of para-form. The specific

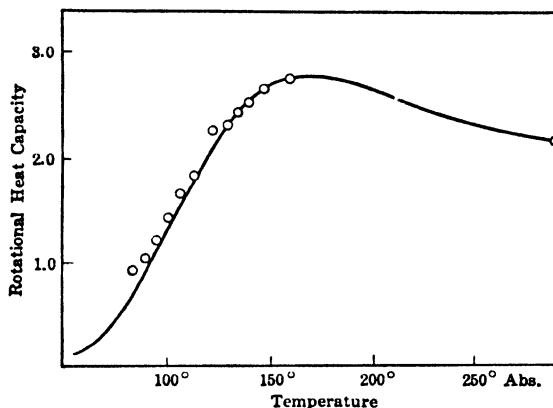


FIG. 54. Rotational heat capacity of parahydrogen

heat should then be calculated by considering the ortho- and para-states as separate gases, their rotational partition functions being written

$$Q_{\text{para}} = \sum_{J=0,2,4,6,\dots} (2J+1)e^{-J(J+1)/kT}$$

and

$$Q_{\text{ortho}} = 3 \sum_{J=1,3,5,7,\dots} (2J+1)e^{-J(J+1)/kT}, \quad (278)$$

and from these C_{para} and C_{ortho} , the rotational molar heat capacities can be calculated by (276). The rotational heat capacity of "normal" hydrogen is then

$$C_{\text{normal}} = \frac{1}{2}C_{\text{para}} + \frac{3}{2}C_{\text{ortho}}, \quad (279)$$

which gives values in almost exact agreement with the experimental data down to -180°C . The heat capacity of pure parahydrogen can also be evaluated from Q_{para} , and the results are in excellent harmony with those observed for this form (Fig. 54). It may be noted that the difference in specific heats of ortho- and para-hydrogen is utilized in the method generally employed to estimate their proportions in a mixture; the heat capacity is related to the heat conductivity (p. 281), and it is the latter which is measured.

High Temperatures.—If the effect of nuclear spin is, for the moment, disregarded, the rotational partition function for any diatomic molecule may be written

$$Q_{\text{rot.}} = \sum_{J=0}^{\infty} (2J+1)e^{-\epsilon_J/kT}. \quad (280)$$

The rotational energy of a rigid molecule is given by $J(J+1)h^2/8\pi^2I$ (p. 566), and if $h^2/8\pi^2IkT$ is small, that is, at relatively high temperatures, especially for molecules with large moments of inertia, the summation may be replaced by integration, with the result

$$Q_{\text{rot.}} = 8\pi^2IkT/h^2. \quad (281)$$

In order to allow for the effect of nuclear spin, this must be multiplied by $\frac{1}{2}(2i+1)^2$ for a molecule with two identical atoms and by $(2i+1)(2i'+1)$ if the atoms are different; if the electronic ground level is a multiplet, the result should be multiplied accordingly. For present purposes, however, these additional factors do not matter since they are constant, and introducing the value of $Q_{\text{rot.}}$ of (281) in (274), it is at once seen that

$$E_{\text{rot.}} = RT, \quad (282)$$

so that the rotational energy of a diatomic molecule is RT per mole, in agreement with the equipartition principle. It can be shown similarly that the rotational energy is the same for any linear molecule, and is $\frac{3}{2}RT$ for one with a nonlinear structure, also in harmony with the principle. The significance of this conclusion is that if the temperature is sufficiently high for a large number of rotational levels to be occupied, the quantum theory and the classical theory give identical results; ordinary room temperatures are generally high enough for this condition to be realized. The difference between the two points of view, however, is that the latter does not allow for the partial excitation of rotational energy levels, whereas the former implies the possibility of various amounts of this energy, i.e., of different rotational levels; the numbers of molecules occupying the different levels change with temperature. The contrast between the classical and quantum treatments is brought out by the fact that the rotational contribution to the molar heat of equilibrium-hydrogen has been calculated to be 4 cal. at 50°K ., and for parahydrogen a value of 2.84 cal. deg^{-1} has been observed at 150°K . (see Fig. 54), whereas the maximum permitted by the classical theory is 1.987 cal. deg^{-1} . The reason for this is that owing to the rapid excitation of rotational energy levels at these temperatures

$d \ln Q_{\text{rot.}}/dT$ exceeds the limiting (high temperature) value of $1/T$ given by (281); such a state of affairs is, of course, quite outside the scope of the equipartition principle.

The energy of each mode of vibration of a molecule, assumed to be simple harmonic, is $h\omega v$ in excess of the zero-point value, where ω is the equilibrium frequency of vibration in wave numbers and v is a quantum number which may be 0 or integral (cf. p. 567). Since the statistical weight of each vibrational level is unity, it follows that the partition function for vibrational energy is

$$Q_{\text{vib.}} = \sum_{v=0}^{\infty} e^{-h\omega v/kT} = (1 - e^{-h\omega/kT})^{-1}. \quad (283)$$

If this value of $Q_{\text{vib.}}$ is inserted in (274), then provided $h\omega/kT$ is small, that is, if the vibration frequency is low or the temperature high, it follows that

$$E_{\text{vib.}} = RT, \quad (284)$$

again in agreement with the equipartition principle. At sufficiently high temperatures, therefore, each mode of vibration contributes RT towards the internal energy. At ordinary temperatures most diatomic molecules are in the lowest vibrational level, and so they appear to be rigid, but when the vibration frequency is low, e.g., for chlorine, then the vibrational energy is of importance even at relatively low temperatures, as seen above. At high temperatures, however, when a number of vibrational levels are occupied in such substances as hydrogen, oxygen and nitric oxide, as well as chlorine and polyatomic molecules, the vibrational internal energy will be the same as that required by the classical principle.

So far the quantum theory treatment has provided a general interpretation of the equipartition principle, but it is important to emphasize that there are cases, such as that of hydrogen, mentioned above, which the principle does not cover. In addition it may be noted that at high temperatures the heat capacities of oxygen and nitric oxide exceed the value of $\frac{1}{2}R$ permitted by the classical method, viz., $\frac{3}{2}R$ for translation, R for vibration, and R for rotation. The explanation is that in both cases the molecules exist in multiplet electronic ground levels, and there is also a possibility, with oxygen, of two excited electronic states in addition to the ground state. Part of the energy of the molecule is, therefore, electronic, so that the total internal energy may rise to a value greater than the maximum permitted by the equipartition principle for translation, vibration and rotation.¹⁸

REFERENCES

- Born, M. "Atomic Physics."
 Fowler, R. H. and Guggenheim, E. A. "Statistical Thermodynamics."
 Jeans, J. H. "Dynamical Theory of Gases"; "Introduction to the Kinetic Theory of Gases."
 Kennard, E. H. "Kinetic Theory of Gases."
 Loeb, L. B. "Kinetic Theory of Gases."
 Taylor, H. S. "Treatise on Physical Chemistry," 2nd. ed, Chaps. III and IV (K. F. Herzfeld and H. M. Smallwood).

1. Stern, *Z. Phys.*, **2**, 49 (1920).
2. Perrin, *Compt. rend.* **146**, 967 (1908); **147**, 530; **149**, 477, 549 (1909); **152**, 1380 (1911); *Ann. Chim. Phys.*, **18**, 5 (1909); "Atoms" (trans. by D. L. Hammick); H. Fletcher, *Phys. Rev.*, **33**, 81 (1911); **4**, 440 (1914); Nordlund, *Z. phys. Chem.*, **87**, 40 (1914); Westgren, *ibid.*, **89**, 63 (1914); *Z. anorg. Chem.*, **93**, 231 (1915); Porter, *Trans. Farad. Soc.*, **18**, 91 (1922); **19**, 1 (1923); Burton, *Proc. Roy. Soc.*, **100A**, 414 (1922); for summaries, see Burton and Smith, "The Physical Properties of Colloids' Solutions," 1938. Chaps. VI, VII and VIII; Virgo, *Science Progress*, **27**, 634 (1933).
3. Costa, Smyth and Compton, *Phys. Rev.*, **30**, 319 (1927); Eldridge, *ibid.*, p. 931; Zartman, *ibid.*, **37**, 383 (1931); Meissner and Scheffers, *Phys. Z.*, **34**, 48 (1933); R. N. Pease, *J. Chem. Ed.*, **16**, 242, 366 (1939); Glasstone, "Theoretical Chemistry," Chap. VII.
4. Knudsen, *Ann. Phys.*, **28**, 999; **29**, 179 (1909); Langmuir, *J. Am. Chem. Soc.*, **35**, 105 (1913); Debiérne, *Compt. rend.*, **150**, 1740 (1910); Weide and Bichowsky, *J. Am. Chem. Soc.*, **48**, 2529 (1926); De Vries and Rodebush, *ibid.*, **49**, 656 (1927); Eyring, *ibid.*, **50**, 2398 (1928).
5. Rankine, *Proc. Roy. Soc.*, **84A**, 181 (1910); **99A**, 331 (1920); *Trans. Farad. Soc.*, **17**, 119 (1922); Rankine and Smith, *Phil. Mag.*, **42**, 601 (1921); Harke, *Proc. Roy. Soc.*, **100A**, 429 (1922); Nasini, *ibid.*, **123A**, 686, 692, 704 (1929); Mack, *J. Am. Chem. Soc.*, **47**, 2468 (1925); Bearden, *Phys. Rev.*, **56**, 1023 (1939); Mason and Maass, *Canad. J. Res.*, **18B**, 128 (1940).
6. Dieterici, *Ann. Phys.*, **11**, 700 (1899); F. H. MacDougall, *J. Am. Chem. Soc.*, **38**, 528 (1916); **39**, 1229 (1917); Jeans, *J. Chem. Soc.*, **123**, 3398 (1923), van der Waals Memorial Lecture; Partington, "Chemical Thermodynamics"; Woolsey, *J. Chem. Ed.*, **16**, 60 (1939); Maron and Turnbull, *J. Am. Chem. Soc.*, **64**, 2195 (1942).
7. Keyes, *Proc. Nat. Acad. Sci.*, **3**, 323 (1917); *J. Am. Chem. Soc.*, **41**, 589 (1919); **60**, 1761 (1938); Beattie and Bridgeman, *ibid.*, **49**, 1665 (1927); **50**, 3133 (1928).
8. Mayer, *et al.*, *J. Chem. Phys.*, **5**, 67, 75 (1937); Lennard-Jones and Devonshire, *Proc. Roy. Soc.*, **163A**, 53 (1937); Collins and Keyes, *J. Phys. Chem.*, **43**, 5 (1939); Hirschfelder and Roseveare, *ibid.*, p. 15; Roper, *ibid.*, **44**, 835 (1940); **45**, 321 (1941); see also Glasstone, ref. 3, Chap. IX.
9. London, *Phys. Z.*, **60**, 491; **63**, 245 (1930); *Z. phys. Chem.*, **B11**, 222 (1930); *Trans. Farad. Soc.*, **33**, 8 (1937); Lennard-Jones, *Proc. Phys. Soc.*, **43**, 461 (1931); Slater and Kirkwood, *Phys. Rev.*, **37**, 686 (1931); Kirkwood, *Phys. Z.*, **33**, 57 (1932); Dushman, *J. Chem. Ed.*, **12**, 581 (1935); **13**, 32 (1936).
10. Masson, *et al.*, *Proc. Roy. Soc.*, **103A**, 524 (1923); **122A**, 283 (1929); **126A**, 248 (1930); Lurie and Gillespie, *J. Am. Chem. Soc.*, **49**, 1146 (1927); Bartlett, *et al.*, *ibid.*, **50**, 1275 (1928); Krichevsky, *ibid.*, **59**, 2733 (1937); Lewis and Randall, "Thermodynamics," 1923, p. 226.
11. Masson, ref. (10); Lennard-Jones and Cook, *Proc. Roy. Soc.*, **115A**, 334 (1927); Keyes, *J. Am. Chem. Soc.*, **49**, 1393 (1927); Beattie, *ibid.*, **51**, 19, (1929).
12. Guye, *Compt. rend.*, **139**, 679 (1904); **141**, 51 (1905); Gray and Burt, *J. Chem. Soc.*, **95**, 1633 (1909); Perman and Davies, *Proc. Roy. Soc.*, **78A**, 28 (1906); Maass, *J. Am. Chem. Soc.*, **40**, 1847 (1918).
13. Whytlaw-Gray, Patterson and Cawood, *Proc. Roy. Soc.*, **134A**, 7 (1931).
14. Berthelot, *Compt. rend.*, **126**, 954 (1898); *J. de phys.*, **8**, 263 (1899); Guye, *J. Chim. Phys.*, **6**, 769 (1908); Baxter, *J. Am. Chem. Soc.*, **44**, 595 (1922); Beattie and Bridgeman, *ibid.*, **50**, 3151 (1928); Cawood and Patterson, *Phil. Trans. Roy. Soc.*, **236A**, 77 (1936); Moles, *et al.*, *Z. anorg. Chem.*, **236**, 225 (1938); *Trans. Farad. Soc.*, **35**, 1439 (1939).
15. Meyer, *Ber.*, **11**, 1867, 2253 (1876); Nernst, *Z. Elek.*, **9**, 622 (1902); von Wartenberg, *Ber.*, **39**, 381 (1906); *Z. anorg. Chem.*, **56**, 320 (1907); Bleier and Kohn, *Monatsh.*, **20**, 505 (1899); Lumsden, *J. Chem. Soc.*, **83**, 342 (1903); Menzies, *J. Am. Chem. Soc.*, **32**, 1624 (1910); Cady, *et al.*, *ibid.*, **63**, 1357 (1941); *J. Phys. Chem.*, **42**, 397 (1938); Wagner, *et al.*, *Ind. Eng. Chem. (Anal.)*, **14**, 129 (1942).

16. Bodenstein, *et al.*, *Z. Elek.*, **15**, 244 (1909); **16**, 961 (1910); Preuner, *Z. physikal. Chem.*, **68**, 129 (1909); Daniels, *J. Am. Chem. Soc.*, **50**, 1115 (1928); see also ref. (4).
17. Holborn and Henning, *Ann. Phys.*, **18**, 739 (1905); **28**, 809 (1907); Scheel and Heuse, *ibid.*, **37**, 79 (1912); **40**, 473 (1913); Partington, *et al.*, *Proc. Roy. Soc.*, **100A**, 27 (1921); **105A**, 225 (1924); *Phil. Mag.*, **43**, 369 (1922); **45**, 416 (1923); *Trans. Farad. Soc.*, **18**, 386 (1923); Partington and Shilling, "The Specific Heats of Gases"; Dixon, *et al.*, *Proc. Roy. Soc.*, **100A**, 1 (1921); **105A**, 199 (1924); Blackett, *et al.*, *ibid.*, **126A**, 319, 333 (1930); Eucken and Weigert, *Z. phys. Chem.*, **B23**, 265 (1933); Bennewitz, *et al.*, *ibid.*, **B39**, 126 (1938); **A186**, 299 (1940); Kistiakowsky, *et al.*, *J. Chem. Phys.*, **6**, 18 (1938); **7**, 281, 289 (1939); **8**, 610 (1940); Telfair, *ibid.*, **10**, 167 (1942).
18. Dennison, *Proc. Roy. Soc.*, **115A**, 483 (1927); Clausius and Hiller, *Z. phys. Chem.*, **B4**, 158 (1929); Giauque, *J. Am. Chem. Soc.*, **52**, 4808, 4816 (1930); Kassel, *Chem. Rev.*, **18**, 277 (1936); Farkas, "Orthohydrogen, Parahydrogen and Heavy Hydrogen"; Fowler and Guggenheim, "Statistical Thermodynamics"; Glasstone, ref. 3, Chaps. IV and VIII.

CHAPTER V

THE SOLID STATE

CRYSTALLINE FORM

Introduction.—The term **solid** is employed in a general sense to describe forms of matter possessing rigidity and a tendency to maintain both their shape and volume; in these respects solids differ from gases and liquids. In recent years, however, it has become the practice to restrict the use of the expression **solid state** to substances which are of a crystalline nature; that is to say, this state is distinguished by a complete regularity of arrangement of the atoms or molecules of which the substance is constituted. In liquids, as well as in some apparently solid materials, e.g., glass and pitch, the distribution of atoms or molecules is not altogether regular; such substances are said to be amorphous, that is, without shape. It will be seen later that even in the amorphous state there is probably some regularity, but it is only partial and possibly transient. Nevertheless, it appears that there may be a gradual transition between the strict regimentation on the one hand and completely random distribution on the other, and a study of the solid state has helped to throw light on the structures of amorphous materials.

Crystal Angles.—The study of crystals has two aspects, namely the examination of the external form and the elucidation of the internal structure; the latter is more important from the chemical point of view, but a knowledge of crystal form is essential before the internal arrangement of the atoms can be fully understood. The size and shape of the crystals of a given compound may vary with the conditions under which crystallization occurs; for example, sodium chloride separates as cubes from a neutral medium, as octahedra if a small amount of urea is in solution, or as tree-like growths if gum arabic is present. In spite of

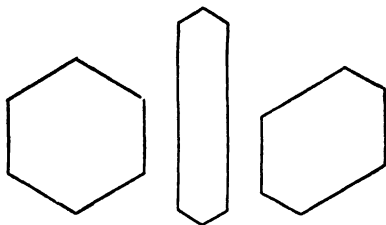


FIG. 55. Crystal habit

this behavior is shown in Fig. 55. The measurement of the angles forms an important part of the study of crystals, and the instrument used for

these differences, the angles between the faces remain constant, as was first realized by N. Steno (1666) for quartz crystals of varying size and shape. The external shape or **habit** of a crystal of a given compound depends on the relative development of the different faces, but the interfacial angles are always the same; a two-dimensional representation of

the purpose is called a **goniometer**. The simplest type, used only for approximate work, is the contact goniometer, in which the angles are measured directly by means of a protractor with a pivoted arm; for accurate determinations, however, the reflection goniometer is invariably employed, the angles between the faces being determined by observation of a reflected beam of light.

Inspection of a crystal, and particularly measurement of its angles, shows that there are always a number of faces which are clearly of the same type, whereas others belong equally definitely to another type. Each set of similar faces is called a **form**, and the faces of any given form make characteristic angles with each other and with other faces. As a consequence of the regular shape of a crystal, faces are found to occur in sets which meet in parallel edges, or would meet in parallel edges if the planes of the faces were extended. Such a set of faces is known as a **zone**; each zone forms a belt completely around the crystal. An axis taken through the center of the crystal parallel to the edges of any zone is a **zone axis**; if the crystal is rotated about this axis, faces belonging to that particular zone will come successively into parallel positions.

Crystal Symmetry.—Another fact which is at once apparent when a crystal is examined is the existence of various types of symmetry; that is to say, by performing a simple geometrical operation, e.g., rotation about an axis, or reflection across a plane, its appearance remains unchanged. Three simple types will be considered here. A body has a **plane of symmetry** when it can be divided by an imaginary plane into two parts, such that one is the exact mirror image of the other; in other words, any line perpendicular to the plane intersects the crystal surface at equal distances on either side. An **axis of symmetry** is a line about which the crystal may be rotated so that it presents exactly the same appearance more than once in the course of a complete revolution. If the self-coincidence occurs twice, that is, the original aspect is repeated as a result of rotation through 180° , the axis is said to be one of two-fold symmetry, or a **diad axis**; other possibilities are three, four and six repetitions of the initial appearance, that is, after rotations of 120° , 90° and 60° , respectively, for three-fold (**triad**), four-fold (**tetrad**) and six-fold (**hexad**) axes of symmetry. In addition, a crystal may possess a **center of symmetry**; this is a point such that any line drawn through it will intersect the surface of the crystal at equal distances on either side. A crystal may have one or more planes, and one or more axes of symmetry, but never more than one center of symmetry; in fact, many crystals are not centro-symmetrical since they develop differently at opposite ends.

The diagrams in Fig. 56 serve to illustrate the significance of the various elements of symmetry for the cube: (a) shows a rectangular plane of symmetry of which there are three, the others being in directions at right angles to the one shown; (b) is a diagonal plane of symmetry, there being a total of six such planes passing diagonally through the cube; (c) indicates one of the three axes of four-fold symmetry, at right angles

to each other; (d) represents an axis of three-fold symmetry, passing through opposite corners, of which the cube possesses four; (e) depicts one of the six axes of two-fold symmetry, emerging from opposite edges; and finally (f) shows the one center of symmetry at the mass center of the cube. A perfect cube has, therefore, a total of twenty-three elements of symmetry; this number is exceptionally large, because of the high degree of symmetry possessed by a cubical body.

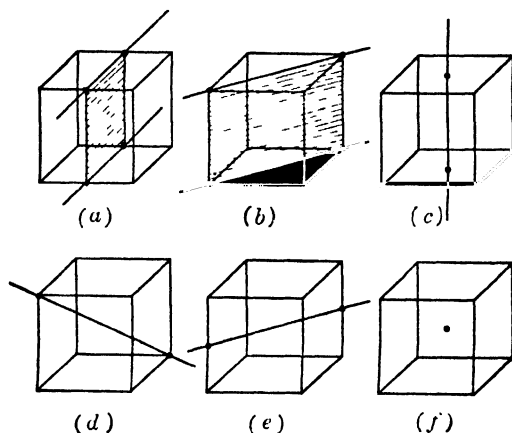


FIG. 56. Elements of symmetry of a cube

Crystallographic Axes and Axial Ratios.—The geometrical form of a crystal is best described in terms of its crystallographic axes; these are generally three, or sometimes four, lines meeting at a point, which are chosen so as to bear a definite relationship to the characteristic features of the crystal. For example, the crystallographic axes may coincide with, or be parallel to, the axes of symmetry, or the edges between principal faces, or be normal to the planes of symmetry. The choice of axes is to some extent arbitrary, although it is found in practice that a certain set of axes leads to a much simpler representation of the crystal than any other selection. Whenever possible the three axes are chosen to be at right angles to each other, but for many crystals this arrangement bears no relationship to the essential structure of the crystal, and the axes must therefore be inclined at different angles. It is now necessary to select a particular plane of the crystal as a **standard** or **unit plane**, in terms of which the faces may be described. The choice is again somewhat arbitrary, although experience generally indicates the best plane for the purpose. One essential property is that it shall cut all three crystallographic axes, the intercepts, a , b and c , being called the **parameters** of

* The intercepts can be calculated from measurements of the interfacial angles of the crystal by a goniometer.

the crystal; in Fig. 57, for example, the axes are OX , OY and OZ , and the unit plane is represented as cutting these axes at A , B and C , respectively, giving intercepts $OA(=a)$, $OB(=b)$ and $OC(=c)$. The lengths a , b and c are purely relative, since their actual values depend on the position of the origin O , but their ratios are constant; the ratios of the intercepts, called the **axial ratios**, are of importance in the description of a crystal. In stating the axial ratios it is the convention to write b as equal to unity; thus $a : b : c = 0.5727 : 1 : 0.7418$ for potassium sulfate. All faces of the same "form" (p. 341) have the same axial ratios. The directions and positions of the axes in Fig. 57 are also fixed by convention, and the angles between them, YOZ , XOZ and XOY , are represented by

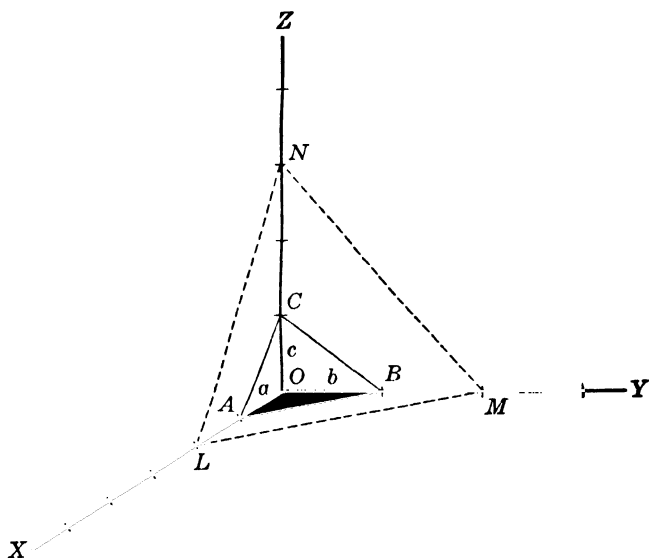


FIG. 57. Crystallographic axes and parameters

α , β and γ respectively. The actual values are quoted, if they are not 90° ; the angles together with the axial ratios, in terms of which the position of any crystal face may be described, are known as the **elements of the crystal**.

Law of Rational Indices.—If the intercepts of any other faces of the crystal on the three axes are measured, they will be found to be represented by la , mb and nc , where l , m and n are *simple whole numbers*, e.g., 1, 2, 3, or occasionally higher; for example, for the face LMN in Fig. 57, the numbers are 2, 2 and 3 respectively. This is an illustration of the **law of rational indices** (R. J. Haüy, 1784), which is one of the fundamental laws of crystal structure; its theoretical basis will be considered later.

In order to represent any particular face in terms of the crystallographic axes and the axial ratios, use is made of the simple relationship

between the latter and the intercepts of that face on the axes. The method employed is that proposed by W. H. Miller (1839); each face is described by three integers, h, k, l , which give the ratio of the intercepts of the unit plane to those of the given face. For the face LMN , for example, the Miller indices, as they are called, are given by the relationship

$$OL : OM : ON = \frac{a}{h} : \frac{b}{k} : \frac{c}{l},$$

the face being described by the symbol (hkl) . Since OL is $2a$, OM is $2b$ and ON is $3c$, it is seen that the ratios are given by

$$OL : OM : ON = \frac{a}{3} : \frac{b}{3} : \frac{c}{2},$$

so that LMN is a (332) face. The Miller indices of a face are thus inversely proportional to the intercepts of that face on the chosen axes. It is obvious that for the unit, or standard, plane ABC , the values of h, k and l are unity, so that it is often described as a (111) face. If the intercept of any plane has a negative value, that is, it cuts the axes when extended in a direction opposite to those shown in Fig. 57, the fact is represented by a bar over the Miller index, e.g., $(\bar{1}\bar{1}1)$ for a face which has intercepts $a, -b$ and c , on the axes OX, OY and OZ respectively. The eight faces of an octahedron are $(111), (1\bar{1}1), (\bar{1}11), (\bar{1}\bar{1}1), (1\bar{1}\bar{1}), (11\bar{1}), (\bar{1}\bar{1}\bar{1})$ and $(\bar{1}\bar{1}\bar{1})$, the three axes being taken at right angles. The faces all have the same axial ratios and consequently belong to one "form," described as $\{111\}$.

When a face cuts only two axes, e.g., OX and OY , and is parallel to the third, the intercept on the OZ axis is then infinite and the corresponding Miller index is 0. If the face is parallel to one of the axial planes, so that it has a finite intercept on one axis only, then the symbol 00 is employed in the Miller index. For a perfect cube, the crystallographic axes being chosen parallel to three edges at right angles and meeting at the center of the cube, each face cuts only one axis and the symbol for the cube form is $\{100\}$; the various faces may be represented by $(100), (010), (001), (\bar{1}00), (0\bar{1}0)$ and $(00\bar{1})$.

Crystallographic Systems.—From considerations of geometry, based on the occurrence of different elements of symmetry passing through a single point, taken in conjunction with the law of rational indices, J. F. C. Hessel (1830) showed that thirty-two different classes of crystal symmetry, or point groups, were possible. These point groups define all the ways in which the elements of crystallographic symmetry can be distributed about a single point in space. Examples of all, with the possible exception of one, are known. The thirty-two classes fall conveniently into seven systems characterized by their axial angles and ratios; six of these systems are represented by three crystallographic axes, but one, the hexagonal system, is conveniently referred to a set of

four axes. The characteristic properties of the various systems, together with the number of symmetry classes in each system, are given in Table 46. It should be noted that the elements of symmetry recorded in the third column represent the *minimum* requirements for each particular system;

TABLE 46. CRYSTALLOGRAPHIC SYSTEMS

System	Crystallographic Elements	Essential Symmetry	Number of Point Groups	Examples
Cubic, or Regular	Three axes at right angles: all equal. $\alpha = \beta = \gamma = 90^\circ$ $a : b : c = 1 : 1 : 1$	4 triad axes; 3 diad, or 3 tetrad axes	5	CaF_2 , ZnS , FeS_2 , Cu_2O , NaClO_3 , NaCl , Diamond, Pb , Hg , Ag , Au
Tetragonal	Three axes at right angles: two equal. $\alpha = \beta = \gamma = 90^\circ$ $a : b : c = 1 : 1 : y$	1 tetrad axis	7	SnO_2 , Sn , TiO_2 , KH_2PO_4 , PbWO_4
Orthorhombic or Rhombic	Three axes at right angles: unequal. $\alpha = \beta = \gamma = 90^\circ$ $a : b : c = x : 1 : y$	3 diad axes, or 1 diad axis and 2 perpendicular planes intersecting in a diad axis	3	PbCO_3 , BaSO_4 , K_2SO_4 , Mg_2SiO_4 , $\alpha\text{-S}$, KNO_3
Monoclinic	Three axes, one pair not at right angles: unequal. $\alpha = \gamma = 90^\circ$, $\beta \neq 90^\circ$ $a : b : c = x : 1 : y$	1 diad axis, or 1 plane	3	$\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$, $\beta\text{-S}$, $\text{K}_2\text{Mg}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$
Trichinic or Anorthic	Three axes not at right angles: unequal. $\alpha, \beta, \gamma \neq 90^\circ$ $a : b : c = x : 1 : y$	No axes or planes	2	$\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, $\text{K}_2\text{Cr}_2\text{O}_7$
Hexagonal	Three axes coplanar at 60° : equal. Fourth axis at right angles to other three. $a_1 : a_2 : a_3 : b$ $= 1 : 1 : 1 : x$	1 hexad axis	7	HgS , Ice, Graphite, Mg , Zn , Cd
Rhombohedral or Trigonal	Three axes equally inclined, not at right angles: all equal. $\alpha = \beta = \gamma \neq 90^\circ$ $a : b : c = 1 : 1 : 1$	1 triad axis	5	Calcite, Magnesite, NaNO_3 , ICl , Quartz, As , Sb , Bi

it is the different variations which determine the possible classes of point groups. The class possessing the highest symmetry possible in its system is called **holohedral**, and has the maximum number of similar faces; in a system of lower class, half of the faces may differ from the other half, and it is then said to be **hemihedral**. In the **tetartohedral** class only one-quarter of the maximum number of faces are developed.

Optical Properties.—The velocity of light, and hence the refractive index, in a crystal belonging to the cubic system, is the same in every direction, as in an amorphous medium, e.g., water or glass; substances having this property are said to be **isotropic**. It arises in cubic crystals because they have the same structure in all directions; crystals of other systems are, however, **anisotropic**, and the velocity of light is not uniform throughout. In general, when a ray of light enters an anisotropic crystal it is split up into two separate components which travel with different velocities; they are consequently refracted to different extents and the two rays follow different paths. This phenomenon, which is well known in connection with Iceland spar, can be detected for all crystals except those in the cubic system; it is called **double refraction** or **birefringence**. There is generally one direction or another along which the crystal is uniform, and it is found that if light is incident on the crystal parallel to a given line, called the **optic axis**, there is only single, and not double, refraction. Crystals of the tetragonal, hexagonal and rhombohedral systems have one optic axis, and are said to be uniaxial, whereas rhombic, monoclinic and triclinic crystals are biaxial, for there are two directions parallel to which light can travel with one definite velocity.

If a convergent beam of white polarized light is passed through a section of a uniaxial crystal cut perpendicular to the optic axis, and examined through a Nicol prism, a characteristic interference pattern will be observed; it consists of a number of concentric circles of colored light surmounted by a dark cross passing through the center of the circles. If the Nicol prism used as analyzer is rotated through 90° the color of the circles will change to their complementary color, and the dark cross will become light. A biaxial crystal gives a different type of interference figure; for its observation it is necessary to cut a section perpendicular to the line bisecting the angle between the two optic axes. The pattern appears as a number of concentric elliptical rings, or lemniscates, crossed by two dark bands or "brushes." The color changes taking place on rotating the analyzer are similar to those for uniaxial crystals. Examination in polarized light is often employed for the identification of crystal systems.

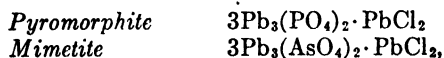
Isomorphism.—In the course of a study of the crystalline phosphates and arsenates of sodium, potassium and ammonium, E. Mitscherlich (1819) observed that a pair of salts of similar constitution, e.g., $\text{KH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$ and $\text{KH}_2\text{AsO}_4 \cdot \text{H}_2\text{O}$, in which the phosphorus atom of one was replaced by an atom of arsenic in the other, had the same crystalline form; according to Mitscherlich, the crystals resembled one another so closely that it was "quite impossible to find any difference." To describe the replacement of one element by another of a related chemical type without producing any appreciable change in crystal form, Mitscherlich used the term **isomorphism** (Greek: *same shape*), and the elements were said

to be **isomorphous**. These expressions are now applied more specifically, however, to the crystalline substances rather than to the elements contained in them; thus, the potassium phosphate and arsenate mentioned above are said to be isomorphous and to display isomorphism.

As a result of his work on the phosphates and arsenates, Mitscherlich concluded that *an equal number of atoms combined in the same manner produce the same crystalline form*, and that *the crystalline form of a compound is independent of the chemical nature of the combined atoms and is determined only by their number and method of combination*. This generalization is known as **Mitscherlich's law of isomorphism**; there are many exceptions to the law stated in this manner, and it is probably more satisfactory to put it in the simpler form, that *substances which are similar in crystalline form and in chemical properties can usually be represented by similar formulae*. For example, Mitscherlich found that sodium selenate was isomorphous with sodium sulfate, and in view of the resemblance between the two salts the former may be assumed to be Na_2SeO_4 . On the other hand, barium permanganate is also isomorphous with these two salts, but its formula is certainly not Ba_2MnO_4 , but $\text{Ba}(\text{MnO}_4)_2$, contrary to the general law.

The possibility of using the suggestion that isomorphous substances have similar formulae for the purpose of determining atomic weights was realized by J. J. Berzelius and applied by him in the table of atomic weights drawn up in 1826; for example, the formula of green chromic oxide was taken as Cr_2O_3 , on account of its isomorphism with Fe_2O_3 and Al_2O_3 . The principle was also used by J. B. A. Dumas (1837) to rectify the atomic weight of silver; this had previously been accepted as 216, and the sulfide was represented as AgS . The isomorphism of the latter with cuprous sulfide, Cu_2S , suggested that the silver compound was Ag_2S , and hence the atomic weight should be half the value given above, i.e., 108. It may be mentioned that lead sulfide, as *galena*, is isomorphous with silver sulfide (*argentite*), although the formulae are PbS and Ag_2S , respectively; this is a serious exception to the law of isomorphism.

An interesting illustration of the employment of the principle of isomorphism is in connection with the chemistry of vanadium. According to Berzelius the oxide of vanadium was VO_3 and the mineral *vanadinite* had the formula $3\text{Pb}_3(\text{VO}_3)_2 \cdot \text{PbCl}_2$. In 1856 C. Rammelsberg showed that this substance was isomorphous with *pyromorphite* and *mimetite*, having the formulae



and consequently H. E. Roscoe (1868) argued that if the law of isomorphism were correct, *vanadinite* should be $3\text{Pb}_3(\text{VO}_4)_2 \cdot \text{PbCl}_2$. This led him to a re-investigation of the compounds of vanadium, and to the discovery that the so-called elementary vanadium of Berzelius was really an oxide, and that his VO_3 was actually V_2O_5 , analogous to N_2O_5 , P_2O_5 , As_2O_5 , etc.

It is true that there are many instances where isomorphous substances contain equal numbers of elements of the same valence. Outstanding

examples are the alums, $M^I M^{III}(SO_4)_2 \cdot 12H_2O$, and the monoclinic double sulfates $M_2^I M^{II}(SO_4)_2 \cdot 6H_2O$ and the corresponding selenates; but there are several isomorphous substances containing elements of different valence types, e.g., KIO_4 , $CaWO_4$ and $KOsO_3N$; $NaNO_3$ and $CaCO_3$; $KClO_4$, $BaSO_4$ and KBF_4 (T. V. Barker, 1912). It is evident that caution must be exercised in the use of the law of isomorphism for chemical purposes. Reference may also be made to another difficulty; sodium and potassium nitrates are not isomorphous, neither are magnesium and strontium carbonates in spite of the similarity of the compounds themselves and of the metallic elements. The explanation of these and other problems will be dealt with later (p. 392).

Variation of Crystal Constants in Isomorphous Substances.—Mitscherlich was originally of the opinion that the crystalline forms of isomorphous substances were indistinguishable, in spite of the fact that W. H. Wollaston (1812) had previously found small differences in the interfacial angles of some isomorphous carbonates, but later he realized that such slight variations are of general occurrence. The term isomorphism may, therefore, be employed to describe substances crystallizing in similar forms and having interfacial angles and axial ratios which do not differ very appreciably from each other.* As an illustration of the type of variation observed, the figures in Table 47 may be considered; they give the interfacial angle and axial ratios for the isomorphous rhombic crystals of aragonite ($CaCO_3$), strontianite ($SrCO_3$), witherite ($BaCO_3$), and cerussite ($PbCO_3$). Similar variations have been found by A. E. H.

TABLE 47. CRYSTALLOGRAPHIC CONSTANTS OF ISOMORPHOUS CARBONATES

	$CaCO_3$	$SrCO_3$	$BaCO_3$	$PbCO_3$
Interfacial Angle	$116^\circ 10'$	$117^\circ 18'$	$117^\circ 48'$	$117^\circ 18'$
Axial Ratios (a:b:c)	0.623:1:0.721	0.609:1:0.724	0.595:1:0.741	0.610:1:0.723

Tutton (1893–1925) in the course of his very accurate and comprehensive measurements of the single sulfates, $M_2^I SO_4$, and of the monoclinic double sulfates, $M_2^I M^{II}(SO_4)_2 \cdot 6H_2O$, and the corresponding selenates.

Definition of Isomorphism.—It is probable that Mitscherlich meant the word isomorphism to apply to substances of similar composition, for it is to such compounds that his law applies, and it is in this sense, too; that later workers have employed the term. A much wider significance is, however, now finding general acceptance; the description “isomorphous” is used, in accordance with its Greek origin, for substances of closely similar crystalline form, e.g., $NaNO_3$ and $CaCO_3$ (*calcite*), irrespective of their chemical nature.

Tests for Isomorphism.—If a solution of two isomorphous salts, e.g., phosphate and arsenate, or a mixture of alums, is allowed to crystallize, the solid is homogeneous but contains *both* substances, the proportions depending on the relative amounts in the liquid. Homogeneous solids of

* In the cubic system the interfacial angles are exactly 90° , and the axial ratios are all unity, so that isomorphous crystals must be identical in form.

this kind are sometimes called **mixed crystals**, a misleading translation of the word "mischkristalle," proposed by H. W. B. Roozeboom (1899); the expression **mix-crystals** is better, but on the whole the term **solid solution**, used in another connection (p. 649) by J. H. van't Hoff (1890), is preferable. The solid is a true homogeneous solution of one substance in another, and has many of the properties of solutions. Just as some liquids are completely miscible with others, so there are completely miscible solids giving a continuous series of solid solutions; similarly, partially miscible liquids have their analogue in partially miscible solids. It was realized by Mitscherlich that isomorphous substances form solid solutions, and J. W. Retgers (1889) considered that the property of forming solid solutions was an important criterion of isomorphism. Retgers postulated that strict isomorphism is accompanied by exact additivity of physical properties, e.g., specific volume, of the constituents in the solid solution, but this test is probably too stringent and is not generally applicable. It is true that many isomorphous substances do form solid solutions, either when crystallization takes place from solution or from the molten mixture, but it appears that appreciable solubility in the solid state occurs only when the two substances have almost equal molecular volumes.

It should be pointed out that not all substances yielding solid solutions are necessarily isomorphous; it is sometimes possible for the molecules of one substance to enter the crystal of another without producing any considerable change of form. For example, sodium chlorate, normally cubic, will dissolve in excess of tetragonal silver chlorate and *vice versa*; the crystal form of the solid solution is determined by the substance present in largest amount.

If a very sparingly soluble substance is isomorphous with another which is readily soluble, e.g., calcium carbonate and sodium nitrate, the test of the formation of mix-crystals is obviously difficult or impossible to apply. It has been suggested (T. V. Barker, 1906) that in such circumstances the formation of **parallel growths** should be observed. It has been long known (M. L. Frankenheim, 1836) that if a drop of sodium nitrate solution is allowed to evaporate on a fresh cleavage face of calcite, small crystals of the former are oriented so as to be very closely parallel to the edges of the calcite cleavage. Potassium perchlorate and permanganate have been shown to form parallel growths on *barytes* (BaSO_4), and are undoubtedly isomorphous with it; the salts of rubidium and cesium have crystallographic constants very close to those of potassium, but they do not behave in the same manner on barytes. This difference is attributed to appreciable differences in the molecular volumes.

J. L. Gay-Lussac (1816) observed that a crystal of potash alum will continue to grow when placed in a saturated solution of ammonium alum, and this ability to form **overgrowths** has been suggested as another criterion of isomorphism (H. Kopp, 1879). Many instances of the production of overgrowths, e.g., simple sulfates of the transition metals, alums,

monoclinic double sulfates, etc., are known. A consequence of this property of isomorphous substances is that each is able to induce crystallization from a saturated solution of the other; this may even occur when one of the substances is sparingly soluble, for it is stated that calcite can facilitate the separation of anhydrous sodium nitrate from solution. It is probable here also that, apart from resemblance in crystalline form, approximate equality of molecular volumes is the determining factor, and so the ultimate test for isomorphism must be similarity of the geometrical constants of the crystals. If the molecular volumes are not very different then the properties mentioned above will be manifest, but otherwise substances with almost identical crystalline form may not show them. As will be seen later, a better understanding of isomorphism has resulted from the investigation of crystal structure by X-ray methods (p. 392).

Polymorphism.—Many substances exist in more than one crystalline form; this phenomenon is known as **polymorphism**. The particular form adopted depends on the conditions of crystallization, e.g., temperature and pressure; the possibility of mutual conversion will be considered in Chapter VI in a discussion of change of state. Polymorphism occurs in substances of various types, element and compound, inorganic and organic; among elements, e.g., carbon, sulfur, phosphorus, etc., it is called **allotropy**. A substance existing in two different modifications is said to be dimorphic; if there are three different forms it is trimorphic, and so on. If each of the two forms of a polymorphic substance is isomorphous with a form of another polymorphic substance the phenomenon is known as **isopolymorphism**; for example, arsenic and antimony trioxides are isodimorphic, whereas tin dioxide and titanium dioxide are isotrimorphic. Isodimorphism occurs among the sulfates $M^{II}SO_4 \cdot 7H_2O$, where M^{II} is magnesium, zinc, nickel, cobalt or manganese; these substances exist in both rhombic and monoclinic modifications. The formation of solid solutions apparently occurs over a limited range of concentration for each crystalline state.

THE INTERNAL STRUCTURE OF CRYSTALS

The Space Lattice.—The regularity of crystal growth and the fact that a calcite crystal could be broken up by cleavage into a number of small rhombs having the shape of the original crystal, led the Abbé R. J. Haüy (1784) to suggest that a crystal is built up of a number of very small units, each having a shape related to that of the crystal as a whole. This view has developed into the concept of the **space lattice**, which is based on the accepted regularity of the internal structure of substances in the crystalline state. Consider, in the first place, a regular two-dimensional design, such as that on a wall-paper; it is evident that it is constructed by the repetition of a definite unit pattern at regular intervals. A particular point can be chosen in one of these patterns and the positions of all similar points in the other patterns marked; these are seen to form

a regular network, and if parallel lines are drawn in two directions through these points the wall-paper will be divided into a series of cells each containing the same pattern. It is immaterial which particular point in the design is chosen in the first place; the network of identical points will always appear the same for the given wall-paper. A crystal is a regular three-dimensional design, and identical points form a three-dimensional network of cells each representing the unit from which the whole crystal is constructed. This regular arrangement of points is called the space lattice; each point in the lattice has exactly the same environment as any similar point. If the crystal is that of a monatomic element, e.g., a metal, then there will be only one space lattice, since there is only one kind of atom. The space lattice of a compound is made up of the interpenetrating lattices of the constituent atoms; in some cases the atom (or ion) lattices are the same, e.g., NaCl , but in others they may be different, e.g., CaF_2 . The symmetry of the combined lattices determines the symmetry of the crystal.

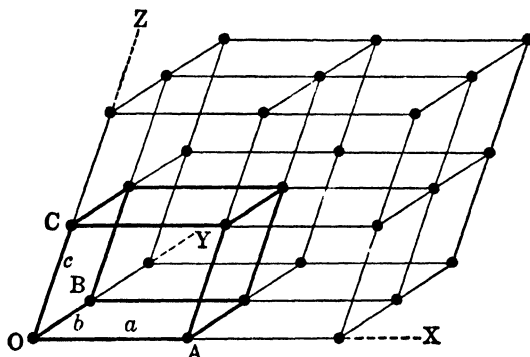


FIG. 58. Space lattice

The points in Fig. 58 may be taken as an example of a space lattice: * for convenience the points may be joined by lines so as to provide three axes whereby the positions of the points can be described. The directions of the lines are arbitrary but there are generally, as in the case of the geometry of the whole crystal, certain directions which give a clearer idea of the lattice structure. In Fig. 58 the directions OX , OY and OZ have been chosen as the axes, and as a result the regular repetition of points within the lattice becomes evident. With these axes the *unit cell* is at once seen as a parallelepiped, one of which, with edges OA , OB and OC , is indicated by thick lines in the figure. This is, of course, not the only way in which a unit cell could have been chosen; others having different shapes, but the same volume, are possible but less convenient. The distances a , b and c , giving the lengths of the edges of the chosen unit

* It should be noted that the *points* constitute the space lattice, and not the lines joining them.

cell, are called its **primitive translations**.^{*} Attention may be called to the fact that in Fig. 58 there is only one point to each unit cell; although the cell has eight corners, with a point at each, it must be remembered that eight cells meet at a corner and the point there is shared equally between each. In other forms of space lattice, e.g., those of the face-centered or body-centered type, there may be more than one point per unit cell.

Lattice Planes.—The points in a space lattice can clearly be arranged in an indefinite number of ways in a series of parallel and equidistant planes; these are the **lattice planes**, or **net-planes**. The faces of the complete crystal are parallel to these net-planes, the most common faces being those corresponding to planes containing the largest number of points, i.e., with the highest **reticular density**. To illustrate this statement it is simplest to consider, in the first place, a two-dimensional lattice, as in Fig. 59; the directions OX and OY are chosen as the axes of reference.

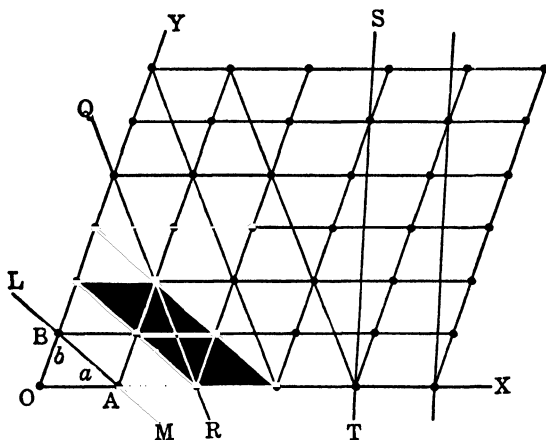


FIG. 59. Lattice planes

The primitive translations of each unit cell are a and b , along OX and OY , respectively. It may now be supposed that the lattice is extended in a third direction, OZ , perpendicular to the plane of the paper, so that a line of points in Fig. 59 represents a section through a perpendicular plane of points. The planes which contain the largest number of points in a given area are obviously those parallel to OX and OY ; the faces parallel to these planes will occur most readily in the complete crystal. The Miller system of indices is employed to describe the planes of points in

^{*} The primitive translations of the unit cell are, as one might expect, related to the axial ratios of the complete crystal; they are either identical or a simple multiple of one another, depending on the choice of the unit plane in the latter case. For example, for *beryl* the axial ratio $a:c$ is 1:0.4989, whereas the dimensions of the unit cell indicate a ratio of 1:0.9956 for the corresponding primitive translations. The latter ratio is almost exactly half the former.

the space lattice in an exactly analogous manner to its use in connection with crystal faces. Planes parallel to OX would consequently have the symbol (010) , since they are also parallel to the OZ axis, and their intercept on the OY axis is equal to the primitive translation; similarly, planes parallel to OY are (100) . It is probable, therefore, that the most common faces will be of the form $\{100\}$.

The set of planes parallel to LM , which have the symbol (110) , also have a relatively high density of points and so the corresponding $\{110\}$ form will also occur in the crystal. As the indices become higher, e.g., planes parallel to QR , i.e., (210) , the number of points per unit area decreases rapidly and the analogous faces are found less frequently in the crystal. The plane of which TS is the section, and those parallel to it, contain such a small number of points that the appearance of the appropriate face in the crystal is very improbable. A face of this kind, if it did occur, would obviously have one high index, viz., (510) ; hence faces with large indices are infrequent. This is in agreement with the law of rational indices, the fundamental basis of which is thus explained by means of the concept of the space lattice.

Spacing of Lattice Planes: Cubic Lattice.—The distance between successive lattice planes of the same type is of importance in the application of X-ray methods to the study of crystals, and a simple example, namely that of a cubic lattice, will be considered to indicate the nature of the results. In Fig. 60 a unit cube is shown, it being assumed that there is a structural unit at each corner; $BDHG$ is a (100) plane, $AEDB$ is a (110) plane, and ABC is a (111) plane. If a is the side of the unit cube, then it is obvious that this represents the separation of successive (100) planes, as well as of other planes of a similar type, e.g., (010) , such as $CDHE$. The next (110) plane to $AEDB$ in the space lattice will be parallel to it and will pass through GH , and simple geometrical treatment shows the distance between two successive planes of this type to be $a/\sqrt{2}$. In addition to the (111) plane ABC , there is another, namely GED , and their perpendicular separation is $a/\sqrt{3}$; this same spacing applies to all the (111) planes of a simple cubic lattice. The separations of (100) , (110) and (111) planes of such a lattice are, therefore, in the ratio of $1:1/\sqrt{2}:1/\sqrt{3}$. For lattices of other types the spacings may be different, but it is a simple problem to evaluate them in any particular case.

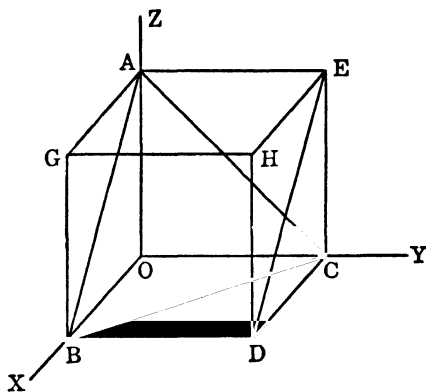


FIG. 60. Unit cube

Types of Lattices.—By means of arguments based purely on geometry, A. Bravais (1848) was able to show that only fourteen kinds of simple space lattice were possible. In other words, there are only fourteen different ways in which similar points can be arranged in a regular or continuous three-dimensional order. The symmetry of these lattices corresponds to the seven crystallographic systems, described above, and their descriptions are given in Table 48. The three types of cubic space lattice are shown in Fig. 61.

TABLE 48. CLASSIFICATION OF SPACE LATTICES

Symmetry Type	Lattice Type
Cubic	Simple cube; face-centered cube; body-centered cube.
Tetragonal	Tetragonal prism; body-centered tetragonal prism.
Orthorhombic	Rectangular prism; body-centered rectangular prism; rhombic prism; body-centered rhombic prism.
Monoclinic	Monoclinic parallelepiped; monoclinic one-face centered.
Triclinic	Triclinic parallelepiped.
Hexagonal	Hexagonal prism.
Rhombohedral	Rhombohedron.

Space Groups.—When considering the external features of crystals it was seen (p. 344) that three kinds of symmetry elements, viz., plane, axis and center of symmetry, exist, and that there are thirty-two possible combinations, or point groups, of elements representing the symmetry of the whole crystal.

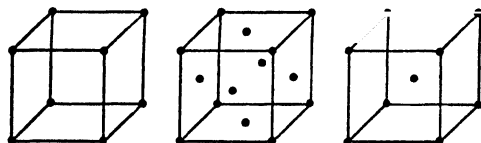


FIG. 61. Cubic lattices

In the complete space lattice, because of the regular repetition of the pattern in three directions in space, two *additional* symmetry elements, viz., **glide planes** and **screw axes**, become possible. Suppose Fig. 62 represents the projection of a portion of a space lattice, in which all the points are the same, the numbers (1), (2), (3) and (4) indicating different levels of the points in the lattice; this is actually the diamond lattice, all the points representing carbon atoms. In order to bring (1), in the center, into the position occupied by (2), it must first be rotated 90° about the vertical axis through A, and then lowered; this operation will bring (2) to (3), (3) to (4), and (4) to a point immediately below (1) in the next layer. If this operation about equivalent axes were performed throughout the whole lattice its *appearance* would be unchanged. The vertical line through A is a four-fold screw axis, since coincidence occurs four times in a complete rotation. Another symmetry operation is to reflect (1) across the plane GG' and then to move it forward and downward so as to occupy position (2); at the same time (2) would go down to (3), (3) would be transferred to (4), and so on. If performed throughout the lattice the appearance of the whole would again be unaltered. The plane GG' is, therefore, an element of symmetry and is called a glide plane.

If in Fig. 62 the points (1) and (3) are different from (2) and (4), e.g., zinc blende, then the four-fold axis about *A* no longer exists, although the glide planes are the same. It is evident, therefore, that because of the difference in the nature of the points the same space lattice can give rise to a different combination of symmetry elements; this is manifested in zinc blende crystals having a lower symmetry than diamond. By purely geometrical methods it has been found that the application of the possible symmetry elements to the points of a space lattice leads to 230 different arrangements, known as **space groups**. Each space group represents a combination of one or more space lattices, always subject to the essential requirement that the aspect from any one point is exactly the same as that from any other point of the same kind. The particular space group to which a crystal belongs is determined essentially by the arrangement with respect to each other of the atoms or ions constituting the compound. The space groups, which thus take into account the internal structure, fall into the thirty-two point groups for the symmetry of the *external* form (p. 344), and consequently into the seven crystallographic systems; the distribution is as shown in Table 49. Special notations have

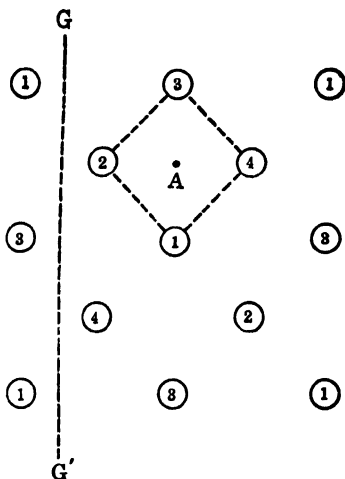


FIG. 62. Glide planes and screw axes

TABLE 49. DISTRIBUTION OF SPACE GROUPS AMONG THE CRYSTALLOGRAPHIC SYSTEMS

System	Number of Space Groups
Cubic	36
Tetragonal	68
Orthorhombic	59
Monoclinic	13
Triclinic	2
Hexagonal	22
Trigonal or Rhombohedral	30

been devised to represent the different space groups; for an account of these the reader must be referred to works on crystallography.

APPLICATION OF X-RAYS

Diffraction of X-Rays.—The classification of crystals described above based on the internal structure was developed by L. Sohncke (1879), E. Fedoroff (1855–90), A. Schönflies (1891), W. Barlow (1894), and others, on theoretical grounds long before any experimental data on the subject were available. A new and significant field of investigation was opened up, however, when it was found that X-rays could be employed to investigate the interior of a crystal. In 1912, W. Friedrich and P. Knipping,

acting on the suggestion of M. von Laue, passed a beam of inhomogeneous X-rays through a slice of zinc blende crystal and allowed the resulting radiation to fall on a photographic plate. Laue was interested in the nature of X-rays, and he considered that if these rays were in fact electromagnetic waves of short wave length, then a crystal might act as a diffraction grating and produce interference effects since the interatomic spacing, viz., 10^{-8} cm., was of the order of the wave length to be expected. The conjecture turned out to be correct, for Friedrich and Knipping found on their plate after development a definite pattern of spots which could only have been due to diffraction of the X-rays. The important discovery that a crystal could behave as a three-dimensional diffraction grating, led W. L. Bragg (1913) to the use of X-rays for the purpose of studying the internal structure of crystals, a subject which owes its early development to the work of this investigator and of his father, W. H. Bragg.¹

The Bragg Equation.—Although von Laue discussed the theoretical aspect of the diffraction of X-rays by crystals, his treatment is more difficult than that of W. L. Bragg (1912) which will be given here in a simplified form. Every atom is able to scatter X-rays to an extent dependent on the number of external electrons, and every atom-bearing plane in a crystal, that is every net-plane or lattice plane (p. 252), behaves to X-rays just as does a line in a diffraction grating to visible light. The positions of the spectra when diffracted by a line grating depend on the distance between successive lines; similarly, the nature of the X-ray diffraction is determined by the spacing between successive planes. Suppose the lines AA , BB , CC , etc., in Fig. 63 represent a number of *identical* net-planes

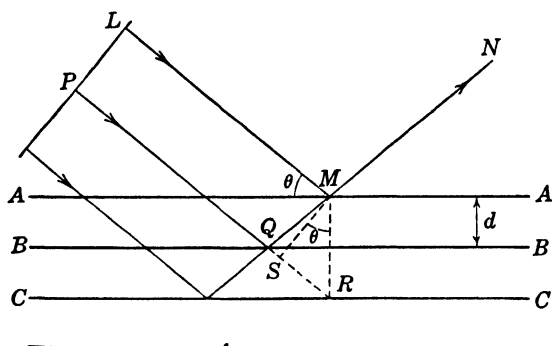


FIG. 63. X-ray reflection from equidistant planes

in a space lattice, the points of which are the atoms constituting the crystal. A parallel beam of X-rays impinges on the crystal so that the glancing angle is θ as shown; part of the beam, e.g., LM , will be reflected at M along MN . On the other hand, some of the rays, e.g., PQ , will penetrate the crystal and be reflected along QN by the atoms in the BB plane; similarly, reflections will occur at other planes. If the path LMN

differs from PQN by a whole number of wave lengths, then the two beams, from L and P , will reinforce one another after reflection and a strong diffracted beam will result. The condition for reinforcement is found in the following manner: from M draw MR perpendicular to the reflecting planes, and MS perpendicular to QR . The difference in the paths LMN and PQN is obviously equal to $QM - QS$, and since $QM = QR$, it is equal to SR . Since the angle SMR is the same as the glancing angle θ , it is seen that SR is given by $2d \sin \theta$, where d is the distance between two successive net planes. For a maximum reflection this must be an integral number (n) of wave lengths (λ); hence the condition, which applies also to rays reflected from other planes, is given by

$$n\lambda = 2d \sin \theta, \quad (1)$$

which is the **Bragg equation**. For a given set of lattice planes and homogeneous X-rays, d and λ are fixed; hence the extent of diffraction will depend on θ , the glancing angle. As θ is increased gradually a series of positions are found, corresponding to $n = 1, 2, 3, 4$, etc., at which maximum reflection occurs, separated by regions in which the various diffracted rays are not in phase and so cancel each other to a great extent. By means of a crystal it is, therefore, possible to obtain an X-ray "spectrum"; the diffraction maxima are called first-, second-, third-, etc., order, according to the value of n . It will be seen that from this spectrum, with the aid of (1), it should be possible to determine the spacing d between successive lattice planes. Each set of planes is capable of producing an X-ray pattern; to identify the reflections a method of indexing is employed similar to that used for lattice planes and crystal faces. The first order reflection from a (hkl) plane is described as hkl , omitting the parentheses, the second order is represented by $2h, 2k, 2l$, and so on; for example, the third order reflection from the (110) plane would be called 330.

In the treatment given above it was assumed that the set of reflecting planes were identical, that is, they contained the same atoms or groups of atoms. In Fig. 64 one set of planes is represented by full lines AA, BB , etc., separated by a distance d , and another set by $A'A', B'B'$, etc., x being the distance between AA and $A'A'$, and between BB and $B'B'$, etc. If the difference of path between the rays LM and RSM is to be a whole number of wave lengths ($n\lambda$), so that they may enhance one another, then for the same glancing angle θ the difference in the paths LM and PQ will be $xn\lambda/d$. If this is also an integral number of wave lengths, then the waves reflected from the $A'A'$, and similar planes, will be in phase with those reflected from BB and with the original rays, so that the corresponding spectra will be particularly strong. On the other hand, if $xn\lambda/d$ is an odd number of half wave lengths, the rays reflected from $A'A'$ will oppose those from BB and the spectra will be weakened; the extent of the decrease depends on the relative reflecting powers of the atoms occupying the different planes. If x is equal to $d/2$, that is $A'A'$ is half

way between AA and BB , then opposition will occur when n is 1, 3, 5, etc.; that is, the odd order spectra will be weak. In the case that x is $d/4$, the weakening is found in the second and sixth orders, and so on. It should be evident, therefore, that information concerning the arrangement of the planes of different atoms in the space lattice can be obtained from a study of the intensity of the X-ray spectra.

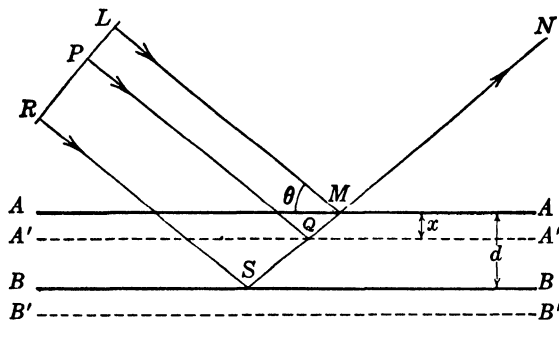


FIG. 64. X-ray reflection from dissimilar planes

Determination of Space Lattice and Space Group.—It has been seen that different spacings and arrangements of reflecting planes lead to spectra of different kinds, and now attention must be paid to the part played by the type of lattice. Consider a lattice containing a number of atoms; the primitive translations of the unit cell along the axes are a , b and c , and choosing a convenient point as origin, let x , y and z be the coordinates of an atom in the directions of the same axes. Suppose a beam of X-rays is reflected in turn from the (hkl) plane passing through the point x , y , z , and then from a parallel plane passing through the origin; it can be shown that the phase difference ϕ between the two wave-trains is given by

$$\phi = 2\pi n(hx/a + ky/b + lz/c), \quad (2)$$

where n is the order of the spectrum. If the coordinates x , y and z are expressed as fractions u , v and w , respectively, of the translations a , b and c , then

$$\phi = 2\pi n(hu + kv + lw). \quad (3)$$

If a unit structure, composed of several atoms, is imagined to be grouped round each point of the space lattice, there will be an (hkl) plane for each atom for which the phase difference for reflection is given by (3) with the appropriate coordinates. In order to determine the resultant amplitude of the wave reflected from the *whole unit*, it is necessary to know the amplitude of the wave for each plane of similar atoms and to combine this with the respective phase difference. The scattering for each plane depends on the number of atoms, which may be taken as constant for all planes of the same type, and on the mean amplitude of the wave scattered by the particular atoms occupying that plane; the latter quantity (f) is called the **atomic scattering factor**. By the rules

of the combination of simple harmonic waves it is found that the resultant amplitude of the hkl reflections, written $F(hkl)$, and called the **structure amplitude**, is given in terms of the square of its "absolute" value by

$$|F(hkl)|^2 = (f_A \cos \phi_A + f_B \cos \phi_B + \dots)^2 + (f_A \sin \phi_A + f_B \sin \phi_B + \dots)^2, \quad (4)$$

where A, B, C, etc., refer to the different atoms in the unit cell, the f 's giving the scattering factors, and the ϕ 's the phase differences according to (3). If the crystal possesses a center of symmetry, as is frequently the case, the sine terms vanish, since the atoms occur in pairs, and for every atom in the lattice with phase ϕ there is a corresponding one with phase $-\phi$, provided the origin is chosen so as to be a center of inversion in the lattice. Equation (4) for the structure amplitude then simplifies to

$$|F(hkl)|^2 = [2(f_A \cos \phi_A + f_B \cos \phi_B + \dots)]^2. \quad (5)$$

The intensity I of the diffracted beam is proportional to $|F(hkl)|^2$, and hence to the expressions on the right hand side of (4) or (5). If the relative intensities of different spectra are determined, it is evident that some information could be obtained concerning the coordinates of the atoms in the structural units.

A simple illustration of the use of (5) is provided by consideration of a face-centered cubic lattice of identical atoms: there is an atom in each corner, making a total of eight, but as they are all shared equally by eight unit lattices there is one corner-atom per unit lattice. Similarly, the six atoms in the centers of the faces are each shared between two lattices, thus leaving three face-centered atoms per unit cell. The points in the unit cell are, therefore, as shown in Fig. 65, marked O, L, M, N; if the point O is chosen as origin the coordinates u, v, w , expressed as fractions of the primitive translations a, b, c , are 000 , $0\frac{1}{2}\frac{1}{2}$, $\frac{1}{2}0\frac{1}{2}$, and $\frac{1}{2}\frac{1}{2}0$, respectively. By means of (3) and (5) it follows that

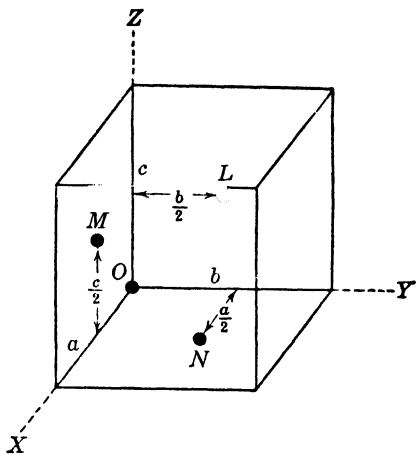


FIG. 65. Face-centered cubic lattice

$$F(hkl) = f[\cos 0 + \cos 2\pi(\frac{1}{2}k + \frac{1}{2}l) + \cos 2\pi(\frac{1}{2}h + \frac{1}{2}l) + \cos 2\pi(\frac{1}{2}h + \frac{1}{2}k)], \quad (6)$$

the sine terms being zero, since the point O taken as origin is a center of inversion. This equation can be rearranged into

$$F(hkl) = 4f \cos \pi(h + k + l) \cos \frac{1}{2}\pi(k + l) \cos \frac{1}{2}\pi(h + l) \cos \frac{1}{2}\pi(h + k), \quad (7)$$

thus giving the condition for all possible reflections from a face-centered cubic lattice. Unless $k + l$, $h + l$, and $h + k$ are all even, that is unless h, k and l are all even or all odd, the equation for $F(hkl)$ reduces to zero; it follows, therefore,

that a face-centered cubic lattice gives reflections only when h, k and l are all even or all odd, e.g., 111, 200, 220, 222, etc., but not 100, 110, 210, etc. In a similar manner it is found that for a simple cubic lattice h, k and l can have any values, but for a body-centered cube $h + k + l$ must be even. It is thus possible to determine the type of lattice provided the indices of the reflections can be identified.

It can also be shown, although the subject is too involved to be discussed here, that each of the 230 space groups gives rise to certain definite characteristics in the X-ray spectrum. The results of the theoretical calculations have been tabulated, and once the observed reflections are correctly indexed it is not a difficult matter to find the particular space group upon which the internal structure of the crystal is built.

Atomic Parameters.—It is possible, at least theoretically, from a close examination of the X-ray spectrum produced by a crystal, to identify both its space lattice and space group, and to determine the size of the unit cell. The number of molecules or atoms per cell can be easily calculated from the density of the substance, and there only remains the problem of the arrangement of the atoms or molecules within the cell. If the material under examination is of simple constitution, e.g., an element or a molecule containing only a few atoms, the information already obtained is sufficient, as will be seen shortly, for the complete internal structure of the crystal to be determined. For relatively complex molecules, however, further data are necessary. The symmetry elements of the space group, the sizes of the atoms present in the molecule, and a knowledge of its chemical structure, will often help to reduce the number of alternatives, but there may still remain a few possibilities for the internal structure which satisfy the X-ray diffraction data. In order to fix the positions of all the atoms there are still some distances, or **parameters**, within the lattice not known definitely; the correct values may be obtained, as will be shown below, by measurement of the intensities of the X-ray spectra obtained from different planes.

Intensity of X-Ray Spectra.—Every element has an atomic scattering factor f , as indicated on p. 358, which represents the average amplitude of the X-rays scattered by a single atom of that element in terms of that scattered by a single "classical" electron. The value of f depends both on the wave length λ of the X-rays and on the glancing angle θ , but it is a function of $(\sin \theta)/\lambda$. When this is small, that is, for small glancing angles, f is almost the same as the atomic number of the element, but it falls off as $(\sin \theta)/\lambda$ increases. The scattering factor for any atom is given by the equation

$$f = \int_0^\infty W(r) \frac{\sin \phi}{\phi} dr, \quad (8)$$

where $\phi = 4\pi r(\sin \theta)/\lambda$, and $W(r)dr$ represents the probability of finding an electron, or the average number of electrons, between two shells r and $r + dr$ from the center of the atom. This probability may be determined by means of wave mechanics, and so the scattering factor can be calculated for different values of $(\sin \theta)/\lambda$ for any atom. In some instances the results have been verified experimentally, so that the data obtained by the theoretical treatment,

which are to be found in tables, may be taken as reliable. These values can be inserted in (4) in order to determine the structure amplitude $F(hkl)$ for the scattering from any particular plane (hkl).

For the calculation of the intensity of the X-rays reflected from the planes of a crystal it is necessary to make certain assumptions; the ideal case would be that of a "perfect" crystal, having mathematically perfect faces. Actually crystals are, of course, imperfect, but the extent of imperfection is not known. It is the custom, therefore, to postulate a "mosaic" crystal, that is, one made up of a number of small almost perfect crystalline blocks inclined at small angles to each other. The reflections from the face of such a crystal must naturally vary over a small angle, and so in the measurement the crystal is turned through a small range in the region of reflection, and the total amount of reflected radiation measured. If E is the total reflected energy, ω is the uniform angular velocity of rotation of the crystal, and I is the total incident radiation falling on the crystal face per sec., then the quantity R ,* defined by

$$R = E\omega/I, \quad (9)$$

is called the **integrated reflection**, and is taken as a measure of the intensity of the reflected beam. From theoretical considerations the integrated reflection from an (hkl) plane of a crystal for a glancing angle θ is given, in terms of the structure amplitude, by

$$R(hkl) = \frac{N^2 e^4}{m^2 c^4} |F(hkl)|^2 \frac{\lambda^3}{2\mu \sin^2 \theta} P, \quad (10)$$

where N is the number of unit lattices per cc. of crystal, e is the electronic charge, m the mass of the atom, c the velocity of light, λ the wave length of the X-rays, and μ is the absorption coefficient of the rays in the material of the crystal; the quantity P is the polarization factor, $(1 + \cos^2 2\theta)/2$, which makes allowance for the unpolarized nature of the incident radiation.

Structure by Trial Analysis.—All the factors on the right-hand side of (10) are generally known except the structure amplitude factor $F(hkl)$ for the given plane, but if the integrated reflection $R(hkl)$ is measured experimentally, it is at once possible to calculate the former quantity. According to (4) the structure amplitude is determined by the known atomic scattering factors and the phase differences ϕ , which depend on the parameters x , y and z , that is, on the positions of the atoms in the unit cell. As a general rule there are too many parameters for their direct evaluation to be possible, but the reversal of this procedure can be carried out with satisfactory results in many cases. The method of "trial analysis" is frequently employed; each of the possible atomic arrangements in space, based on the considerations already mentioned (p. 355), is taken, and using the appropriate coordinates of the atoms, the values of $F(hkl)$ to be expected for reflections of different orders for various planes are determined by calculation [equation (4)]. The structure whose coordinates give results for the structure amplitude in best agreement with those found experimentally from the intensities [equation (10)] is taken as correct. As a rule there is little ambiguity and the positions of the atoms in the space lattice are obtained with some certainty.

* The symbol ρ is generally used for the integrated reflection, but as it is also employed for electron density (p. 362), the change has been made here to avoid confusion.

Structure by Fourier Synthesis.—An alternative procedure, depending on the use of Fourier series, first outlined by W. H. Bragg (1915) and suggested independently by W. Duane (1925), was applied to simple cases by R. J. Havighurst (1925) and A. H. Compton (1926); an important advance was, however, made by W. L. Bragg (1929) who used the method in the investigation of the structures of complex silicates. It has since been frequently employed in the elucidation of the crystal structure of relatively complicated organic molecules.

Since every crystal has a regular periodicity, that is to say, a definite pattern repeats itself regularly throughout the whole crystal, the X-ray scattering matter of which it is constituted is capable of being represented by a complicated, but nevertheless periodic, function in three dimensions. Any quantity which is a periodic function of a single coordinate x can be represented by a simple Fourier series, viz.,

$$F(x) = A_0 + A_1 \cos(2\pi x/a + \alpha_1) + A_2 \cos(4\pi x/a + \alpha_2) + A_3 \cos(6\pi x/a + \alpha_3) + \dots \quad (11)$$

$$= \sum_{n=0}^{\infty} A_n \cos(2\pi n x/a + \alpha_n), \quad (12)$$

where A is the amplitude and α the phase of each cosine term; a is the distance within which the periodic function repeats itself. The density of scattering matter throughout a crystal, that is, the electron density, since electrons are responsible for X-ray scattering, can similarly be expressed by a triple Fourier series in terms of three coordinates x , y and z measured parallel to the edges of the unit cell (a , b and c); thus for a crystal the electron density at any point is represented by

$$\rho(xyz) = \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} A(hkl) \cos(2\pi hx/a + 2\pi ky/b + 2\pi lz/c + \alpha_{hkl}), \quad (13)$$

where the various symbols have the significance already indicated. The amplitude factor $A(hkl)$ is related to the structure amplitude and the volume of the unit cell, and for a crystal possessing a center of inversion which is used as origin, (13) can be written as

$$\rho(xyz) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F(hkl) \cos 2\pi(hx/a + ky/b + lz/c), \quad (14)$$

where V is the volume of the unit cell, and $F(hkl)$ is the structure amplitude. In principle, (14) should permit of the complete solution of the structure of a crystal. Provided the dimensions of the unit cell are known, and $F(hkl)$ can be determined from the observed intensities of the X-ray spectra for a large number of values of h , k and l , the triple Fourier series may be calculated* for different values of the coordinates x , y and z . If this can be done, a three dimensional model can be constructed showing the variation of electron density throughout space. Since the electrons are associated with atoms, the latter must be where the density is greatest; in other words, the coordinates of the atoms are determined. Apart from the arithmetical labor involved, there is a fundamental difficulty in the application of this method: (10) gives the *square*

*Strictly speaking an infinite number of terms are required, but as the intensity of the scattered X-radiation falls off rapidly at the higher orders, i.e., large values of h , k or l , a relatively small number of measurements are sometimes adequate.

of the structure amplitude, and consequently although its numerical value may be known its *sign* is uncertain; this is not obtainable without at least an approximate knowledge of the coordinates of the atoms in the crystal. The method generally employed to overcome this difficulty is to make a preliminary analysis of the X-ray diffraction measurements, by the methods described above, and so develop an approximate structure, sometimes only for the most important scattering atoms; this is used to determine the signs of the phase constants and then the Fourier synthesis can be applied so as to give more exactly the coordinates in the unit cell. In this manner accurate parameters can be obtained more readily than by the trial analysis procedure, especially for complicated molecules.

A reasonably accurate evaluation of the triple Fourier series [equation (14)] would necessitate an almost impossible number of intensity measurements; the procedure has, therefore, been greatly simplified by supposing the structure to be projected in turn along the directions of the three axes. A double Fourier series is then obtained giving the electron density distribution on the plane of the other two axes; thus for a projection parallel to the *a* axis, i.e., on to the (100) plane,

$$\rho(yz) = \frac{1}{C} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} F(0kl) \cos 2\pi(ky/b + lz/c), \quad (15)$$

where *C* is the cross-sectional area of the cell projected in the same direction.* Analogous equations apply for projections along the other two axes. A reasonably accurate solution is now possible, after a preliminary structure has been used to give the sign of the *F(hkl)* terms, with about one hundred intensity measurements; the resulting reduction of both experimental and arithmetical labor brings the Fourier synthesis within the bounds of practical possibility. The electron density ρ is calculated for different values of the two variable coordinates for each projection, viz., *y* and *z*, in (15), employing the measured intensities of reflection from (0*kl*) planes, and the results are represented in the form of a "contour map," in which positions of equal electron density are joined by a series of contour lines. The diagram thus obtained by J. M. Robertson (1933) for anthracene, for the projection parallel to the *b* axis, is shown in Fig. 66; the carbon atoms are situated at the peaks, the higher ones representing two atoms vertically superimposed when projected. The equivalent structure and the directions of the *a* and *c* axes of the unit cell are also shown; the hydrogen atoms provide such little scattering that their positions must be inferred. It is evident that the molecule of anthracene does not lie flat on the *ac* plane but appears almost edgewise when viewed in the direction of the *b* axis. By plotting similar diagrams for the two other possible projections, a complete and accurate model in three dimensions can be obtained of the molecule of anthracene as it exists in the crystal. It may be noted that workers in the field of X-ray crystallography have developed various devices which simplify the lengthy calculations involved.

In a few instances the Fourier synthesis has been performed without the aid of a preliminary structure; this depends on the availability of two closely related substances, forming very similar crystals, differing only in the respect that an important scattering center has been introduced or altered. The alums

* It will be noted that (0*kl*) represents planes in one zone, i.e. parallel to one common direction (see p. 341).

provide an illustration of the application of this principle (J. M. Cork, 1927; C. A. Beevers and H. Lipson, 1935); the alkali metal in a cesium alum, because of its higher atomic number, makes a larger *positive* contribution to the structure amplitude $F(hkl)$ than does the equivalent atom in a potassium alum. If the value of $F(hkl)$ for a given reflection hkl is increased numerically by the substitution of cesium for potassium then $F(hkl)$ is positive, but if the numerical value is decreased by the replacement then $F(hkl)$ must be negative in the potassium compound. The same principle has been used to determine the crystal structure of phthalocyanine, the comparison of structure amplitudes for various hkl -reflections being made between the parent substance and its nickel derivative. In this manner J. M. Robertson (1936) made the first direct

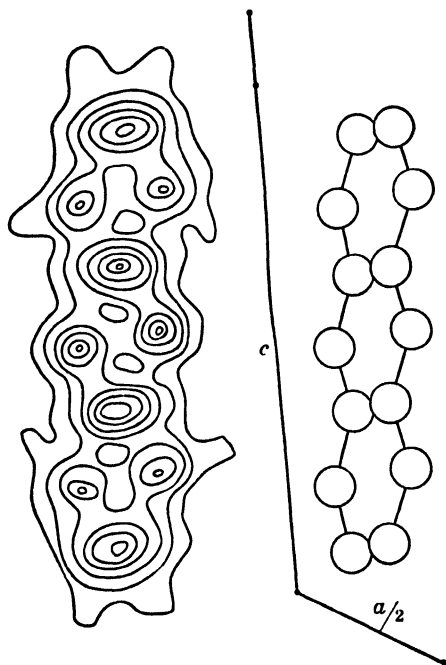


FIG. 66. Electron density and projection of anthracene (Robertson)

complete determination of the structure of a complex organic molecule by X-ray methods.³

The Patterson-Harker Synthesis.—An alternative use of Fourier series in connection with the elucidation of crystal structure has been proposed by A. L. Patterson (1935) and modified by P. Harker (1936). If the intensity of the X-ray reflection, which is proportional to the *square* of $F(hkl)$ and is consequently positive, replaces the structure amplitude in the triple Fourier series equation (14), a Patterson series

$$P(xyz) = \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} \sum_{-\infty}^{\infty} |F(hkl)|^2 \cos(hx/a + ky/b + lz/c) \quad (16)$$

is obtained; this function $P(xyz)$ has maximum values at vector distances from

the origin representing the separations between points of maximum electron density in the crystal.* In other words, the positions of maxima in the Patterson series, relative to the origin of the coordinates x, y, z , give the distances between pairs of atoms in the space lattice. The difficulty in evaluating the triple Fourier series could be overcome by projection on to a plane, as described above, but the results are not satisfactory and an alternative method is adopted. If the crystal has a two-fold symmetry axis, e.g., the a axis, then (16) reduces to the double series

$$P(0yz) = \sum_k \sum_l \cos(ky/b + lz/c) \left\{ \sum_h |F(hkl)|^2 \right\}, \quad (17)$$

the summations being assumed, as usual, to be theoretically from $-\infty$ to $+\infty$; similar simplifications can be made if the structure has other elements of symmetry. The values of $P(0yz)$ for various coordinates y and z are calculated, using the measured reflection intensities which are now not restricted to faces in one zone, i.e., $|F(hkl)|^2$ and not $|F(0kl)|^2$. If the results are plotted as a contour diagram the distance from the origin to the peaks gives the most important interatomic distances; some uncertainty may be involved in the correct allotment of these distances, but a careful consideration of the form of the diagram provides the essential clues. The structure obtained in this way may be refined by the ordinary Fourier synthesis, since the signs of the $F(hkl)$ terms can now be determined; this procedure has been adopted in the direct elucidation of the structure of pentaerythritol (E. G. Cox, *et al.*, 1937).⁴

EXPERIMENTAL METHODS AND RESULTS

Several methods have been developed for the study of X-ray diffraction for the purpose of investigating the internal structure of crystals; although the Laue photograph method was historically the first to be used for this purpose (W. L. Bragg, 1913), the application of the X-ray spectrometer, designed by W. H. Bragg (1913), will be described first because of the greater ease of interpreting the results for simple molecules.

The X-Ray Spectrometer.—A beam of X-rays of definite wave length, obtained from an X-ray tube with a molybdenum or copper anticathode A , passes through a slit B , and falls on to a known face of the crystal C mounted on a rotating table, the position of which can be read on the vernier D (Fig. 67). The radiation reflected from the crystal passes through another slit and into the ionization chamber E containing methyl bromide vapor, the walls being insulated and connected to a source of high potential; the entry of X-rays causes ionization so that there is a flow of current measured on an electrometer. The arm carry-

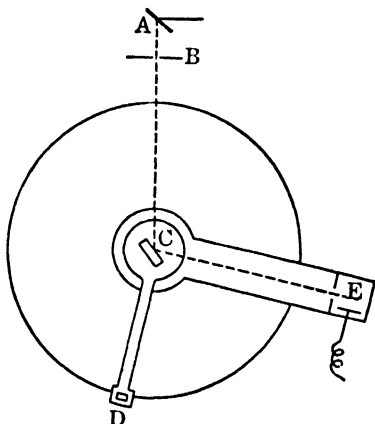


FIG. 67. X-ray spectrometer (Bragg)

ing E rotates about the axis of the instrument at twice the rate of rotation of the table carrying the crystal, so that the reflected ray always enters the ionization chamber for all positions of the crystal. Starting with a small glancing angle between the incident X-rays and the crystal, the value is increased in stages by rotating the table, and readings of the electrometer deflection are taken at each stage; these readings give a measure of the intensity of the reflected X-rays. The glancing angles (θ) for which sharp reflection maxima are obtained are those satisfying the Bragg equation, $2d \sin \theta = n\lambda$, where d is the spacing of the lattice planes parallel to the face of the crystal exposed to the X-rays (cf. p. 357). The procedure is repeated for all important planes of the crystal, and if necessary artificial faces must be cut, e.g., a (111) plane in a perfect cube, which are not present in the natural crystal. The spectrometer method requires the use of a relatively large crystal with well defined faces, and since reflections of several orders must be examined for a number of faces the total labor involved is large; this is compensated for, however, by the comparatively easy interpretation of the results. At the present time the method is chiefly employed for the quantitative determination of the intensities of X-ray reflections, the importance of which has been seen, and for the purpose of checking special points concerning which there may be doubt.⁵

The use of the ionization spectrometer may be illustrated with some simple examples. The (100) face of a sodium chloride crystal was found to give reflection maxima, for X-rays (K_α) from a palladium anticathode, at glancing angles of 5.9° , 11.85° and 18.15° ; if these angles represent first, second and third order reflections, respectively, then according to the Bragg equation their sines should be in the ratio of 1:2:3, and the actual values, viz., 0.1028, 0.2054 and 0.3115, are in excellent agreement with expectation. The Bragg equation may be written in the form

$$d = n\lambda / (2 \sin \theta), \quad (18)$$

and putting $n = 1$ and $\sin \theta = 0.1028$, it follows that d_{100} , the distance between successive *identical* (100) planes, is

$$d_{100} = 4.85\lambda,$$

where λ is the wave length of the X-rays employed.

The first order reflections from (110) and (111) faces of sodium chloride, also with the palladium anticathode, were observed to be at 8.4° and 5.2° , respectively, so that by (18) the ratios of the spacings parallel to the three principal planes are

$$\begin{aligned} d_{100}:d_{110}:d_{111} &= \frac{1}{\sin 5.9^\circ} : \frac{1}{\sin 8.4^\circ} : \frac{1}{\sin 5.2^\circ} \\ &= 9.731 : 6.844 : 11.04 \\ &= 1 : 0.704 : 1.136. \end{aligned}$$

This ratio is almost identical with $1:1/\sqrt{2}:2/\sqrt{3}$, which gives the relative spacings for the unit cell of a face-centered cubic lattice. It appears, therefore, that this type of lattice forms the basis for the structure of the sodium chloride crystal; since the substance belongs to the cubic system the conclusion is reasonable. Further examination of the reflections from the (111) planes shows that whereas the second and fourth orders are normal, the first and third are unusually weak; as seen on p. 357, such a weakening is to be expected if midway between two net planes of the same kind there appears one of a different type. These results are completely interpreted by the sodium chloride lattice shown in Fig. 68, where

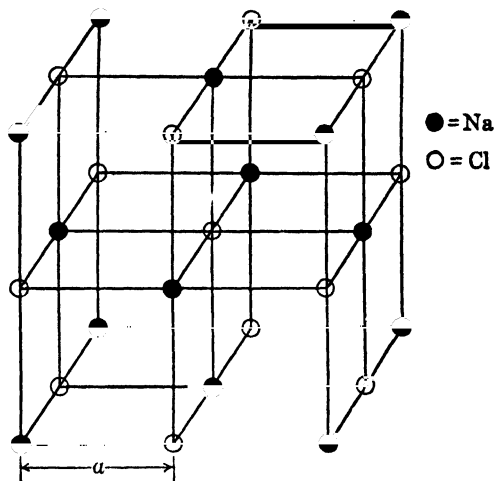


FIG. 68. Sodium chloride lattice

the black circles represent sodium and the white circles chlorine atoms or ions, or *vice versa*.^{*} Only a limited portion of the lattice is shown, but it extends, of course, throughout the whole of the crystal. It is seen that each chlorine unit is surrounded by six sodium units, and each sodium by six chlorine units; nevertheless, as will be found later, there is a balance throughout the whole crystal. It will be noted that there are no sodium chloride *molecules* as such in the lattice; in fact the crystal itself may be regarded as one giant molecule made up of individual sodium and chlorine units.

A careful study of Fig. 68 shows that each set of units forms a face-centered cubic lattice, and that successive (100) and (110) planes contain both sodium and chlorine points. If a is the length of the edge of each

^{*} The sodium and chlorine units are represented in Fig. 68 by relatively small circles, for purposes of clarity; in the crystal, however, the atoms, or ions, are probably almost in contact leaving little free space. This important point should always be remembered when studying conventional lattice diagrams.

small cube, then $d_{100} = a$ and $d_{110} = a/\sqrt{2}$, the ratio being equal to that found experimentally. When the (111) planes are examined it is evident that successive layers contain either sodium or chlorine points only, a plane of one type being exactly midway between a pair of the other kind. The separation of identical planes in the lattice is $2a/\sqrt{3}$, in agreement with the observed result.

With potassium chloride, which is isomorphous with sodium chloride, the first order reflection from the (100) plane occurred at 5.3° , so that

$$\frac{d_{100} \text{ KCl}}{d_{100} \text{ NaCl}} = \frac{\sin 5.9^\circ}{\sin 5.3^\circ} = 1.11.$$

Since the volume of a small lattice cube is $(d_{100})^3$ it follows that if potassium and sodium chlorides have the same crystal structure, the quantity $(1.11)^3 = 1.37$ should give the ratio of the molecular volumes of the two salts; the experimental value 1.39 suggests that the space lattices of the chlorides are in fact the same. An examination of the first order spectrum from the (100), (110) and (111) planes of potassium chloride shows, however, that the ratio of the spacings is represented by

$$d_{100}:d_{110}:d_{111} = 1:1/\sqrt{2}:1/\sqrt{3},$$

which is that required for a *simple* cubic lattice (p. 353). This apparent discrepancy is readily explained when it is realized that the X-ray scattering factor for an atom, at small glancing angles, is equal to the number of extranuclear electrons, i.e., to the atomic number. Since potassium and chlorine have atomic numbers differing only by two units, and potassium and chlorine ions have actually equal numbers of electrons, it is evident that either as atoms or ions their scattering powers will be almost identical. If the space lattice of potassium chloride is, as is very probable, the same as that of sodium chloride (Fig. 68), it is clear that if all the lattice points have equal scattering factors the whole system will behave to X-rays as if it consisted of a lattice of small simple cubes. In sodium chloride the scattering factors of the sodium and chlorine units are different, and so the true structure as two interpenetrating face-centered cubes is apparent.

Another simple, but interesting, case of crystal analysis is that of zinc blende, also belonging to the cubic system. The positions of the first order reflections indicated a face-centered cubic lattice, and although the intensities of the (110) reflections were normal, those from the (100) and (111) planes were not. The odd order (100) reflections were weak, indicating equidistant alternating (100) planes, like the (111) planes of sodium chloride, whereas for the (111) reflection from zinc blende the second order was exceptionally weak, a result suggesting that successive planes were different, with a spacing of dissimilar planes one-quarter that of the similar ones (see p. 358). The structure showing these require-

ments is reproduced in Fig. 69; both zinc and sulfur atoms lie on face-centered cubic lattices, although this is not very obvious, as far as the sulfur is concerned, because of the limited portion of the lattice given in the diagram. The sulfur atoms are at the centers of alternate small cubes, each being surrounded by four zinc atoms at the corners of a regular tetrahedron; similarly, each zinc atom is surrounded tetrahedrally by four sulfur atoms. From the figure it is seen that alternate (100) planes consist of zinc or sulfur points only, the planes of the latter being midway between two of the former. Successive (110) planes contain both types of atom, but the (111) planes also show alternation, the separation between successive zinc and sulfur planes being one-quarter of that between two zinc planes.

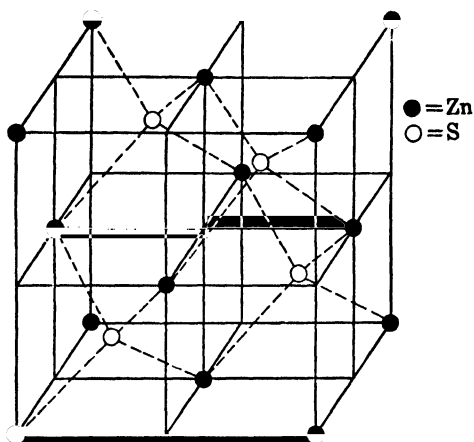


FIG. 69. Zinc sulfide (zinc blende) lattice

The crystals described above are so simple that qualitative data concerning the intensities of the lines are sufficient for the elucidation of the structure, but this is not always the case. In many instances quantitative measurements have been made and used to fix uncertain parameters by the method of trial analysis (p. 361).

Lattice Dimensions.—It has been seen on p. 366 that the distance between successive (100) planes, and consequently of the other planes, in a crystal lattice may be expressed in terms of the wave length of the scattered X-rays; if the latter is known the lattice spacings can be calculated. The earliest measurements of the wave lengths of X-rays were made by W. H. Bragg and W. L. Bragg (1913) utilizing the knowledge gained concerning the structure of crystals. The principle involved may be illustrated by reference to Fig. 68, the unit cell of the space lattice of sodium chloride. In the diagram there are fourteen sodium-bearing points; one at each of the eight corners and one in the middle of each of the six faces. The eight points at the corners are shared equally by eight cubes meeting

at each corner, so there is effectually only one per unit cell; similarly the points on the six faces are each shared by two cubes giving an average of three for each unit. There are thus four sodium atoms associated with the portion of the lattice shown. Similarly, of the thirteen chlorine points, the one at the center belongs exclusively to the unit cell, but the other twelve are each shared between four cubes meeting at the edges; there are thus four chlorine atoms belonging to the unit cell. Four molecules of sodium chloride are thus associated with the large cube in Fig. 68, its length being $2a$ and the volume consequently $8a^3$. If M is the molecular weight of sodium chloride and N the Avogadro number, $4M/N$ is the mass of four single molecules, and if ρ is the density of the salt the mean volume

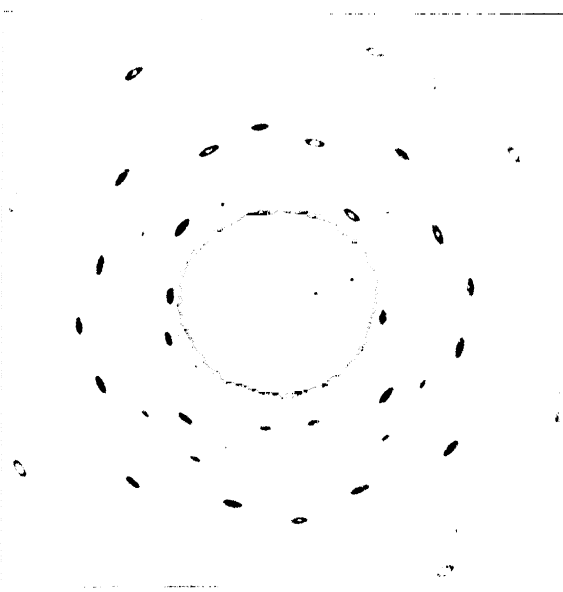


Fig. 70. Laue photograph of magnesium oxide

occupied will be $4M/N\rho$, which should be equal to $8a^3$, the volume of the unit cell; hence,

$$a = \sqrt[3]{M/2N\rho}. \quad (19)$$

Inserting the values $M = 58.5$, $N = 6.02 \times 10^{23}$, and $\rho = .217$, the value of a is found to be 2.82×10^{-8} cm., and since d_{100} is equal to a (p. 368), this also gives the spacing of the (100) planes. For sodium chloride, $d_{100} = 4.85\lambda$, and so λ , the wave length of the K_α X-rays from palladium (p. 366), is 0.581×10^{-8} cm. The results obtained for the wave lengths of X-rays by the use of a crystal as a diffraction grating have been confirmed by A. H. Compton and R. L. Doan (1925) who employed an ordinary ruled line grating and very small glancing angles (see p. 13).

As the wave lengths of the X-rays are known, the dimensions of a unit cell of any crystal can be calculated from the diffraction observations for various crystal planes with the aid of the Bragg equation. By reversing the calculation described above for sodium chloride it is possible to determine the number of molecules associated with a unit cell of the space lattice. This information is often of great assistance in deducing the crystal structure from X-ray data.⁶

The Method of the Laue Photograph.—A narrow beam of X-rays from a tungsten anticathode, consisting of radiations covering a range of wave lengths, is passed through a small crystal, or thin slice of crystal, and the resulting diffracted beam allowed to fall on a photographic plate; development of the latter shows the Laue pattern (Fig. 70), a regular arrangement of spots characteristic of the crystal. The observations are often made with the specimen mounted in such a manner that the incident rays are parallel to an important axis of the crystal; the interpretation of the photograph is then simplified.

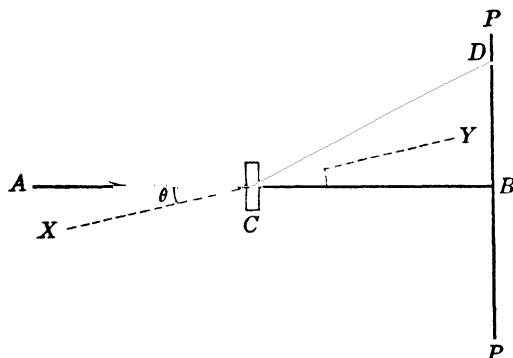


Fig. 71. Principle of the Laue photograph

Suppose AB in Fig. 71 represents the incident inhomogeneous X-rays passing through the crystal C , and XY is the direction of a reflecting plane in the crystal, perpendicular to the plane of the paper; one of the components of the rays of appropriate wave length will then be diffracted toward D in the plane of the paper. If the reflecting plane is imagined to be rotated about XY , the diffracted ray will trace out an ellipse on the photographic plate PP ; the points on a given ellipse, therefore, are produced by planes parallel to the same zone axis, e.g., XY . Since the crystal has a limited number of planes parallel to a given zone axis the Laue photograph shows an equivalent number of well defined spots; each of the spots on one ellipse is produced by planes having identical values for two, of their three, Miller indices, e.g., 111, 121, 131, 141. Every lattice plane in the crystal is parallel to two zone axes, and so the Laue spots are observed at points where the two ellipses intersect; use is made of these facts in the assignment of the spots to the various reflecting planes. The relative spacing between the planes is estimated by calculating θ in Fig. 71 from the positions of the spots and the distance of the photographic plate from the crystal. The interpretation of a Laue photograph in terms of the space lattice

of the crystal was first made by W. L. Bragg (1913) who employed a "stereographic projection," but later the "gnomonic projection" has been used with considerable success by R. W. G. Wyckoff (1920, *et seq.*). A great deal of information can be obtained from one Laue photograph, but the chief disadvantage of the method lies in the uncertainty of the significance of reflection intensities, mainly because of the inhomogeneous nature of the incident X-rays.

At the present time the Laue method is chiefly employed in a modified form, known as the "back-reflection" procedure, which does not require a thin crystal, as in the transmission method just described. The X-rays pass through a narrow slit in the photographic plate and are reflected back from the crystal on to the plate. The method in this form is used to obtain preliminary information concerning the symmetry of the lattice which is utilized in the oscillating crystal procedure described below.⁷

The Rotating Crystal Method.—The method of using a rotating crystal for X-ray analysis was interpreted by M. Polanyi (1921); in one form or another it is probably used more than any other device for the study of crystal structure. A crystal, which should be small enough not to absorb all the radiation in which it is bathed, is rotated round an axis parallel to one of the crystal axes, and exposed to a beam of homogeneous X-rays from a direction at right angles. For any particular position of the crystal there may be no diffraction, but as it rotates various planes come successively into suitable positions for diffraction to occur and corresponding spots are produced on a photographic plate. The conditions for diffraction are shown in Fig. 72, in which XY is the axis of rotation and A and B represent points on two successive lattice planes. The incident X-rays are shown at LA and MB , and the diffracted rays at AP and BQ . For a diffraction maximum to occur the difference in the path AP and BQ , that is BR , must be equal to $n\lambda$, i.e., a whole number of wave lengths. The value of n will depend on the angle BAR (ϕ), and a series of directions of diffraction will be obtained for increasing values of ϕ , as n increases from zero upward. All lattice planes having the same spacing (AB) in the direction parallel to the axis of rotation, will produce diffraction spots lying on a more or less horizontal line (Fig. 73); such lines are known as **layer lines**, and a series is obtained corresponding to the different n values.

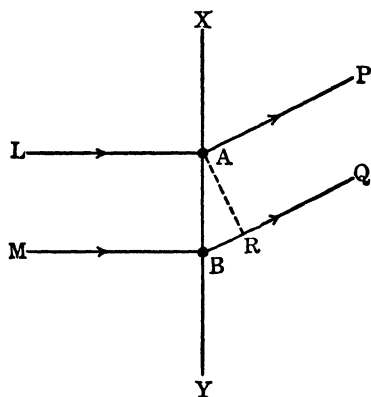


FIG. 72. Principle of the rotation photograph

From the vertical distance between the layer lines, and that from the crystal to the photographic plate, it is possible to calculate ϕ , provided the wave length of the X-rays is known, and hence to determine the distance AB , i.e., the spacing of the planes. In addition to the layer lines it is seen that the spots fall on a set of curves in a transverse direction; these are called **row lines**, and their positions are of importance in the interpretation of the results. By taking three photographs, with the crystal rotating about each of three axes, the lattice spacings and the size of the unit cell can be rapidly deduced. The exact identification of the spots with various orders

of spectra from different planes is not always a simple matter, but graphical devices have been introduced to expedite the work.

Oscillating Crystal Method.—In order to facilitate the interpretation of the results, a modified form of the rotating crystal method has been developed in which the specimen is oscillated through an angle of 10° to 20° ; this limits the number of reflecting positions exposed to the incident X-rays. In the apparatus described by K. Weissenberg (1924), the movement of a cylindrical photographic film is correlated with the oscillations of the crystal, thus facilitating identification of the spots. At the same time, by surrounding the crystal with a sleeve having a slot which is set so as to pass one layer line only, the number of spots on the photograph is decreased, thus simplifying the interpretation. Various forms of this procedure are being widely employed, particularly when an accurate knowledge of the intensities of the reflections is required (cf. p. 360).⁸

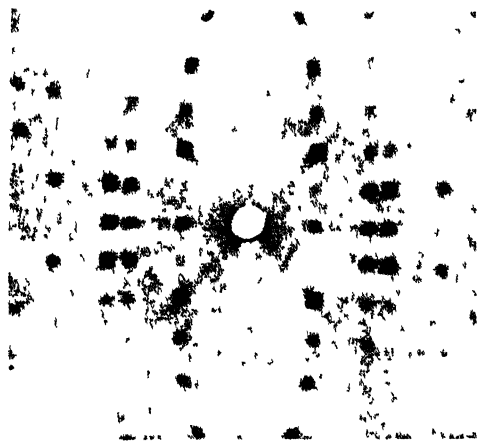


Fig 73. X-ray rotation photograph of benzil

The Powder Method.—The methods described above all require single crystals, even though they need not be very large; the procedure devised independently by P. Debye and P. Scherrer (1916) and A. W. Hull (1917) can be carried through, however, with a fine crystalline powder. A narrow beam of uniform X-rays falls on the finely powdered substance to be examined, and the diffracted rays are passed on to a strip of film which almost completely surrounds the specimen. In the fine powder the crystals are oriented in all directions, and so a large number will have their lattice planes in the correct direction for the Bragg equation, for the diffraction maxima, to be applicable. Suppose θ is the glancing angle necessary for the first order reflection from a given, e.g., (100), plane; all crystalline particles whose (100) planes make an angle θ with the incident beam will, therefore, produce reflections. Since this condition can be satisfied by a series of orientations parallel to a common axis, the diffracted rays will lie on a circular cone as shown in Fig. 74, I; AB is the direction of the incident X-rays, C is the powder, and DCD' represents the cone of rays for the first order diffraction from the (100) planes of a large number of

crystals. Similarly, some crystals will be oriented in such a way as to produce a circular cone of rays ECE' , as the second order reflection from (100) planes, and so on for other planes and for higher orders. If a strip of photographic film is placed at FF' it will cut the cones in such a way as to produce two arcs (Fig. 74, II); each pair gives the position of a reflection of a definite order from a particular plane. The magnitude of the glancing angle can be calculated if the distances from the arc to the position of the undeflected beam, e.g., DB , and from the crystal to the photographic film are known. for the semi-vertical angle of the cone is 2θ ; this argument will be made clear by analogy with Fig. 71. If the indices of the various diffraction arcs can be determined, the lattice spacings can then be calculated provided the wave length of the X-rays is known.

The spacings of cubic crystals can all be expressed in terms of one parameter, the side of the unit cube, and the identification of the lines in the powder photograph is relatively simple; hexagonal, rhombohedral and tetragonal crystals have only two variables, viz., a length and an axial ratio, and the indexing of the lines is not very difficult. The work is greatly facilitated by the graphs prepared by A. W. Hull and W. P. Davey (1921). For crystals of lower symmetry the number of lines is so large that the powder photograph is too

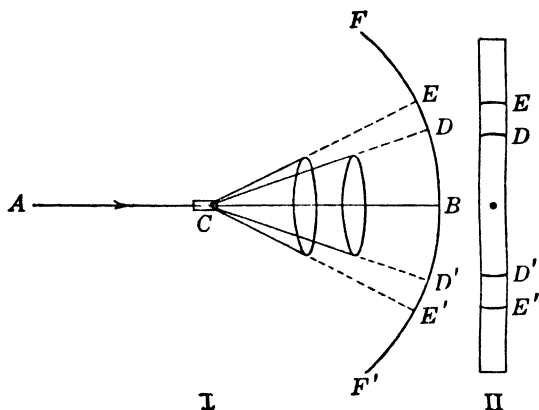


FIG. 74. Principle of the powder photograph

complicated for satisfactory identification of the lines to be possible. The powder method has been mainly employed in the study of metals and alloys; single crystals are difficult to obtain and it is fortunate, therefore, that many of these substances belong to the cubic or hexagonal systems. The X-ray powder photograph is a characteristic property of a crystalline material and so it has been employed as a method of identification and comparison.⁹

Measurement of Intensities.—As already indicated, the ionization spectrometer is most suitable for the purpose of measuring the intensities of X-ray spectra; although photometric determinations have been made of the intensities of the spots on a photographic plate, or film, in the rotating crystal method, the results must be standardized by comparison with the absolute data given by the ionization method. The common practice is to employ a standard crystal, e.g., rock salt, for purposes of comparison. A method has been described (J. M. Robertson, 1934) whereby the crystal under examination

and a standard crystal are alternately placed in the X-ray beam for identical periods of time and the spectra recorded on the same moving film; the intensities may then be compared by photometric examination of the blackening produced.

RESULTS OF CRYSTAL STRUCTURE MEASUREMENTS

A very large number of substances have now been examined by the X-ray method and many of the results obtained are of considerable importance to chemistry. It will only be possible to give here an outline of what appear to be the most interesting conclusions.

Elements: Nonmetals.—The crystal structures of many elements which are solid at ordinary temperatures, as well as some of those solidifying at low temperatures, have been studied; two main types of structures, corresponding to metallic and nonmetallic behavior, have been identified. A few elements, as might be expected, exhibit intermediate properties. Nonmetallic elements are characterized by the fact that each atom is joined by a definite covalent bond to one or more adjoining atoms, the actual number being equal to the usual valence of the element. It follows, therefore, that the crystal of a nonmetallic element is constructed in such a manner as to allow every atom to complete its octet of outermost electrons (p. 99). The number of nearest neighbors to any particular atom in a space lattice is called the **coordination number** of the atom, and for the nonmetallic elements in the fourth, fifth, sixth and seventh groups of the periodic system it is equal to $8 - n$, where n is the group number.

Carbon, in the form of diamond, belonging to the cubic system, was one of the earliest crystals to be investigated by X-rays (W. H. Bragg and W. L. Bragg, 1913); it has the same space lattice as zinc blende (Fig. 69), except that all the atoms are identical. The essential point of this structure, mentioned on p. 369, is that every atom is surrounded by four other atoms situated at the corners of a regular tetrahedron; this may be seen more clearly in Fig. 75, the positions of the centers of the carbon atoms

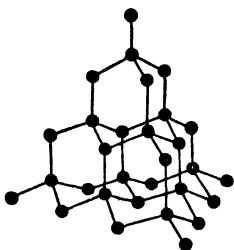


FIG. 75. Space lattice of diamond

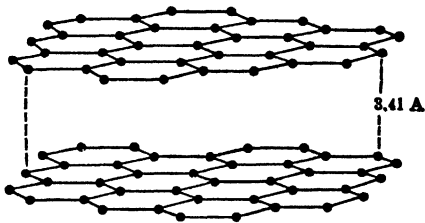


FIG. 76. Space lattice of graphite

being represented without the usual rectangular framework.* The distance between the centers of two adjoining atoms is 1.54 Å., which corre-

* It may be emphasized, again, that the points in the figure represent the centers of the atoms which are probably almost in contact; there is consequently little free space in the crystal (compare footnote, p. 367).

sponds very closely to the distance between two carbon atoms attached to each other by a single covalent bond in aliphatic organic compounds. This agreement, together with the fact that each carbon atom in diamond has four others situated round it at the corners of a regular tetrahedron, suggests that every atom is joined to four others by covalent linkages. A diamond crystal is thus to be regarded as one large carbon molecule, a type of structure for which the term **macromolecule** has been proposed. It is probable that the hardness of diamond, and similar substances, e.g., SiC , is to be ascribed to the strength of the chemical bonding of the atoms and its uniformity in all directions throughout the crystal. It will be noted from Fig. 75 that the carbon atoms form a series of hexagonal "puckered" rings similar to those in cyclohexane, where the carbon atoms are joined by single linkages.

The physical properties of the other allotropic form of carbon, viz., graphite, of the hexagonal system, are quite different from those of diamond, and these differences are found to correspond to important changes in the internal structure of the crystal, as is evident from Fig. 76. The carbon atoms are arranged in flat sheets 3.41 Å. apart; this distance is too large to correspond to a chemical bond, and so each carbon atom is attached to only three others, instead of four as in diamond. In each layer the carbon atoms are seen to be in flat hexagons, the distance between the centers of two adjacent atoms being 1.42 Å.; this arrangement corresponds very closely to the six-membered rings found in benzene, naphthalene and other aromatic hydrocarbons. The carbon atoms in each sheet in the graphite crystal are thus joined by covalent bonds which, as a result of resonance (p. 112), partake, as in benzene, of both single and double bond character. Solids consisting of sheets of atoms, not necessarily all in one plane, extending throughout the whole crystal and separated from one another by a distance which is too large for chemical bonding, are said to have **layer lattices**. They generally have the property of forming flakes readily; the separate sheets in the crystal are held together by relatively weak forces, e.g., of the van der Waals type, and so cleavage easily occurs between them. The use of graphite as a lubricant depends on this ability of one plane of atoms to slide easily over another. It may be recorded that powder photographs of the so-called amorphous forms of carbon show them to consist mainly of very small crystals of graphite.

The elements silicon, germanium and gray tin of the fourth group all have structures of the nonmetallic type similar to that of diamond; the crystals are thus macromolecules in which each atom is joined to four others disposed tetrahedrally. In the fifth group the atoms are trivalent, requiring only three additional electrons to complete their octets, and hence in the crystals it is found that each atom is attached by covalent bonds to three others. Violet and black phosphorus,* arsenic, antimony

* As white phosphorus changes to red under the action of X-rays its structure cannot be determined with certainty.

and bismuth, the last three being rhombohedral, all form lattices of the layer type; the atoms are, however, not all in one plane but consist of puckered six-membered rings, shown in plan in Fig. 77. The atoms of the elements selenium and tellurium, belonging to the next group of the periodic classification, have six valence electrons and their octets can be completed by sharing pairs with two other atoms. In harmony with the views expressed above, concerning the structure of crystals of the nonmetallic elements, each atom is found to be attached by covalent bonds to two others in the form of a spiral chain (Fig. 78). The whole crystal is made up of a series of such chains held parallel to one another by relatively weak forces. Rhombic sulfur resembles the other elements in the same group in the respect that each atom is joined to two others, but instead of chains the atoms form puckered eight-membered rings; this result is of

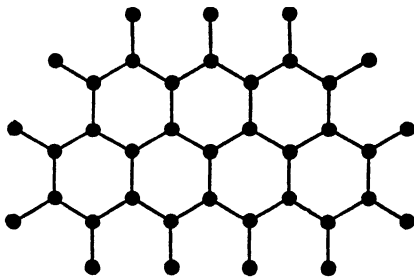


FIG. 77. Space lattice of rhombohedral arsenic, antimony and bismuth

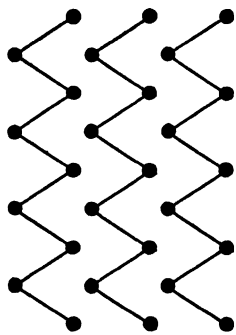


FIG. 78. Space lattice of selenium and tellurium

special interest in view of the existence of sulfur as S_8 molecules in the vapor just above its boiling point. It may be noted that plastic sulfur apparently consists of long chains of sulfur atoms. Iodine is the only halogen whose crystal structure has been determined; the lattice is apparently face-centered rhombohedral, a simple I_2 molecule being situated at each lattice point. This is the first case of the elements so far mentioned, with the exception of rhombic sulfur, in which the normal chemical molecule can be said to exist in the crystal; iodine, therefore, provides an example of a **molecular crystal**. Each iodine atom can be attached to only one other iodine atom, without contravening the octet rule, and so combination in pairs is the only type of union reasonably possible. The formation of covalent bonds thus accounts for the transition from macro-

molecules, through layers and chains, to single molecules as the number of electrons in the valence group increases from four, in carbon, silicon, etc., to seven in iodine.

Metals: The Metallic State.—One of the most important results of X-ray crystallography is the light it has thrown on the metallic state. The great majority of metallic elements crystallize in one of three forms, viz., face-centered cubic, close-packed hexagonal, and body-centered cubic; the first two classes are of special interest as they represent the closest packing of spherical units. Suppose a number of spheres are laid

close together, as shown by the full circles in Fig. 79, and then others are placed in the hollows formed where the spheres meet in the lower layer, indicated by the dotted circles, so as to form a close-packed arrangement. A third layer of the same type may be built up in two alternative ways: either the spheres may be placed in the hollows marked *A*, in which case the spheres in this layer will be exactly over those in the first, or they may be placed in the hollows marked *B*, giving a different type of close-packed arrangement in which the spheres in the fourth layer will correspond with those in the first. On examination, the first type of structure is found to represent hexagonal close-packing, and the second face-centered cubic packing. Since a large proportion of the metallic elements fall into these two classes, it is apparent that the crystals consist of approximately spherical atoms packed together as closely as possible. The body-centered cubic lattice possessed by some metals is also a compact form of packing, but not quite so efficient as that of the other two types.

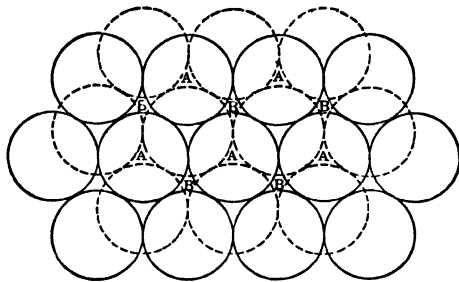


FIG. 79. Close-packed spheres

Mention may be made of a few exceptions to the above generalizations: these are, notably, aluminum, zinc, cadmium, lead and white tin. Zinc and cadmium are close-packed hexagonal, but the distance between the atoms in one direction is greater than the minimum for closest possible packing. Similarly, aluminum and lead have face-centered cubic lattices, but the distances between the atoms exceed the value for the most efficient packing.*

The Metallic Bond.—In both the face-centered cubic and close-packed hexagonal types of structure, each atom is surrounded equally by twelve others, so that the coordination number is twelve. In the body-centered cube the coordination number is lower, namely eight, but still high compared with the values for nonmetallic elements. It is at once evident that the atoms cannot be held together by ordinary covalent bonds resulting from the normal sharing of pairs of electrons, since the number available is insufficient for the purpose. Nevertheless metallic crystals possess considerable strength and it is clear that the forces holding the atoms together are stronger than those of the van der Waals type. It has been

* Gallium is apparently quite exceptional; it crystallizes in the orthorhombic system, and instead of being monatomic like other metals, it forms Ga_2 molecules in the solid which persist in the liquid state.

suggested, therefore, that metals are characterized by the presence of a special type of **metallic bond**, in which a relatively small number of electrons are able, in some manner, to bind together a larger number of atoms or, more correctly, ions. It is probable that the metallic atoms lose some or all of their valence electrons and so become positive ions; the electrons then form a more or less mobile lattice which seems to bind the ions together. The metallic bond has to some extent the character of both covalent and electrovalent linkages, but the properties associated with it are so exceptional as to make it desirable to regard it as a linkage of a quite different type. Nonmetals generally have a high electron affinity and so the atoms tend to appropriate electrons to themselves, and the mobility which is essential to the metallic state cannot arise.

Many of the physical properties of metals can be correlated with the structure just outlined: the high electrical conductivity, for example, is explained by the mobile electron lattice. The ductility and malleability of most metals are attributed to their close-packed structure, for in this way the largest number of "planes of flow" can be realized; the possibility of "flow" in the crystal makes fracture under stress less probable than with nonmetallic substances. White tin crystallizes in the tetragonal system and is not close packed; it is, therefore, less ductile than other metals. It is also a poorer conductor of electricity and hence the metallic bonding is not very marked. As mentioned above, gray tin has the diamond structure and is consequently a nonmetallic solid.

According to L. Pauling (1937 *et seq.*) the metallic bond has essentially a covalent character, involving resonance between a number of structures with one-electron and electron-pair bonds. Increase in the number of such bonds among which resonance can occur should result in greater stability; this is in harmony with the close-packed structure that is so common among the metals, since this gives the highest coordination numbers. An argument in favor of the foregoing view is that transitions from covalent to metallic structures, or the reverse, are relatively common, but transitions from ionic to metallic bonds are rare. For example, in the series of elements, carbon, silicon, germanium and tin, there is a regular transition from covalent to metallic character; further, compounds such as FeS_2 (*pyrite*), usually regarded as covalent, have some metallic properties, e.g., appearance and electrical conductance (cf. pp. 387, 390). The γ -phases of alloys exhibit the reverse behavior, as explained below.

Metallic Alloys.—Fused metals are frequently miscible and on cooling the liquid the two components may either (i) each separate in the pure state; (ii) form a continuous series of solid solutions; (iii) form two solid solutions because of partial miscibility of the solids; or (iv) combine to produce intermetallic compounds (see Chapter X). The first possibility is of no interest from the standpoint of crystal structure, but the others must be considered. X-ray examination has shown that in solid solutions the atoms of one metal replace those of the other at random in the space lattice; provided the two metals have the same lattice structure and their atoms do not differ appreciably in size then, as a general rule, a continuous

series of solid solutions is possible. If one atom is larger than the other, its introduction into the lattice will result in a strain, and a given concentration will be reached when further introduction of foreign atoms will no longer be possible; each element will then have a limited solubility in the other. Sometimes miscibility occurs with two metals that normally have different space lattices; for example, nickel has a face-centered cubic, and chromium a body-centered cubic structure, but nickel will dissolve in solid chromium to the extent of 37 atomic per cent, the alloys being body centered. Any attempt to increase the proportion of nickel, however, results in a change to the face-centered type of lattice characteristic of pure nickel. Although the two metals are completely miscible, the solid solutions do not all have the same structure.

Copper crystallizes in a face-centered cubic form, whereas zinc is close-packed hexagonal; about 32 per cent of zinc will dissolve in copper (α -brass) and about 5 per cent of copper in zinc (η -brass) without change of the respective lattice structures. Between these limits, however, three different solid forms, called β -, γ - and ϵ -phases, respectively, exist; the first of these is body-centered cubic, the second has a complex cubic structure containing fifty-two atoms in the unit cell, and the third has a close-packed hexagonal lattice, but its dimensions differ appreciably from those of zinc. Although the three solids do not necessarily separate out as pure substances, it has been generally accepted, for various reasons, that they are intermetallic compounds of composition CuZn , Cu_2Zn_8 and CuZn_3 , respectively. Further investigation has shown that other pairs of metals form β -, γ - and ϵ -phases with similar crystal lattices, although the complex γ -form may contain from 52 to 416 atoms in the unit cell. The compositions of the compounds formed are, however, not always analogous, but there is nevertheless a regularity of fundamental importance (W. Hume-Rothery, 1926; A. J. Bradley, 1927). If the number of valence electrons in the "compound" is divided by the number of atoms it contains, a definite ratio appears for each type of phase. For the purpose of this calculation the number of valence electrons is counted from the last complete shell; that is to say, it is in general the number in excess of 28, 46 or 78. The elements of the eighth group of the periodic system are assumed to have no free electrons. Some of the results for alloys of different types are given in Table 50. The significance of these ratios was not at first clear, but H. Jones (1934) and others have shown that a satisfactory interpretation may be given to them by means of wave mechanics. It is of interest to mention that the γ -phases are generally brittle substances and poor conductors of electricity; it appears, therefore, that these alloys have some covalent character, thus providing support for the view that the metallic bond is itself a type of covalence.

Although the various phases described above contain two elements in stoichiometrical proportions and their formation is associated with a change in the space lattice, the substances may differ from true compounds in certain respects. In a true compound there is a regular distribution of the constituent atoms throughout the space lattice, but in many of the so-called intermetallic compounds this is not the case. The positions of the atoms in the lattice seem to be determined by considerations of size, although the formulae in the instances cited above are dependent on the number of free electrons associated with the two atoms. It must be emphasized that not all intermetallic com-

TABLE 50. RATIO OF FREE ELECTRONS TO ATOMS IN ALLOYS

	Alloy	Number of Electrons	Number of Atoms	Ratio
<i>β-Phases</i>	CuZn	3	2	3:2
	AgCd	3	2	3:2
	AuZn	3	2	3:2
	Cu ₃ Al	6	4	3:2
	Cu ₃ Sn	9	6	3:2
<i>γ-Phases</i>	Na ₃₁ Pb ₈	63	39	21:13
	Cu ₃ Zn ₈	21	13	21:13
	Cu ₃ Al ₄	21	13	21:13
	Cu ₃₁ Sn ₈	63	39	21:13
	Ni ₃ Zn ₂₁	42	26	21:13
	Ag ₃ Cd ₈	21	13	21:13
<i>ε-Phases</i>	CuZn ₃	7	4	7:4
	Cu ₃ Sn	7	4	7:4
	AgCd ₃	7	4	7:4
	Au ₃ Sn	7	4	7:4

pounds fit into the scheme described above, but only those of a particular type; in some alloys it is probable that there are various covalent bonds, whereas in others the definite atomic ratios are determined by considerations of size only.¹⁰

Order-Disorder Transitions.—The unusual nature of intermetallic compounds is evident from another property, of certain alloys, discovered by means of X-rays. Copper and gold form solid solutions, but examination of the physical properties suggests the existence of two compounds, viz., CuAu and Cu₃Au. If solids of these compositions are properly annealed, the copper and gold atoms are found to occupy definite positions in the lattice. Both copper and gold form face-centered cubic lattices, and in Cu₃Au, for example, the gold atoms appear at the corners of each unit cube and the copper atoms at the centers of the faces. Structures of this type, in which certain regularly chosen positions in the lattice of a given element are occupied by atoms of another element, are called **superlattices** or **superstructures**. If the alloy is not annealed, however, but rapidly cooled from a high temperature, the distribution of copper and gold atoms is quite random, although the structure and size of the unit cell in the lattice remain unchanged. Annealing, therefore, results in the formation of a substance which is apparently a compound, as suggested by the ordered arrangement of the atoms, whereas quenching gives a solid solution, characterized by a random distribution, although there is no observable change in the size and nature of the space lattice. This behavior, which has been called **order-disorder transformation**, has been observed with other pairs of metals, e.g., iron-aluminum. It emphasizes the difficulty, once more, of regarding solids made up of two metallic elements in simple stoichiometric proportions as compounds, for if Cu₃Au were a definite entity it would separate in its own characteristic form irrespective of the rate at which the liquid is cooled. The ordered arrangement presumably occurs because this gives a system of somewhat lower potential energy than the completely random, i.e., disordered, distribution.¹¹

Interstitial Compounds.—The transitional metals, e.g., chromium, manganese, iron and molybdenum, form with the light elements, hydrogen, boron,

carbon and nitrogen, a number of compounds, e.g., Cr_2N , Mn_4C , Fe_2N , Fe_3C , Fe_2B , Mo_2C , etc., possessing metallic characteristics, and differing markedly in this respect from the corresponding compounds of other metals, e.g., NaH , CaC_2 , AlN , which are definitely nonmetallic. By means of X-rays it has been shown that in the substances of the former category the atoms of the light element do not replace those of the transitional metal in the space lattice, but enter the gaps in the lattice producing only a small increase in its dimensions. Systems of this kind have been called **interstitial structures** (G. Hagg, 1929). If the radius of the inserted atom is less than about three-fifths of the other, the compounds have a simple structure and the lattices, like those of most metals, are of the close-packed type; nitrides and some carbides fall into this group. On the other hand, if the inserted atom is larger, in proportion to that of the metallic atom, more complex structures result; examples are the borides of iron, and the carbides of chromium, manganese and the iron-group metals.

INORGANIC COMPOUNDS

Ionic and Covalent Linkages.—It has been seen in connection with the study of the elements that when the lattice coordination number is four or less there is a possibility of the formation of covalent linkages between an atom and its neighbors in the crystal, but this does not occur when the coordination number is larger than four. The same generalization can probably be extended to *simple* inorganic compounds, although for these substances the failure to form covalent linkages leads to the existence of electrovalences and not metallic bonds. In the sodium chloride lattice (Fig. 68) each "atom" has a coordination number of six and a reasonable structure based on covalences is impossible; it is very probable, therefore, in harmony with modern views on the structure of simple salts (p. 98), that the units in the crystal are sodium and chlorine *ions*, which are held together by electrostatic forces. In zinc blende (Fig. 69), on the other hand, the coordination number is four and there is a possibility that each atom may share a pair of electrons with four others, so that the whole molecule is held together by covalent forces. It is by no means certain that the bonds between zinc and sulfur are covalent, since an electrovalent structure, with only Zn^{++} and S^{--} ions, is equally reasonable; it is in fact probable in this, as well as in other, cases that the bonds in the lattice are of an intermediate nature. At the other extreme there are molecular crystals, such as oxygen, hydrogen chloride, iodine, carbon dioxide, stannic iodide, etc., in which the atoms are definitely united by covalent bonds and the individual chemical molecule is the unit of crystal structure. Substances of this nature can be readily identified by X-ray examination.

Atomic and Ionic Radii.—From a consideration of crystals of the two extreme types it has been possible to assess the radii of various elements both in their ionic and atomic forms; the assumption made is that each atom or ion behaves as if it were a sphere of definite size, and in a compound the distance between the centers of two atoms or ions is equal to the sum of their characteristic radii for the particular type of binding. If, however, the linkage is between two atoms with different electro-

negativities, the bond distance is shortened by an amount which increases with this difference. In calculating atomic radii for substances in which the atoms are joined by covalent bonds, the values for a few atoms, such as carbon, whose crystal structure is known, are taken as starting points; from these, others can be determined from measurements on the appropriate compounds. The apparent radius of an atom varies with the multiplicity of the covalent bond and the lattice coordination number; the atomic radii generally tabulated are those for univalent bonding and a coordination number of four, since these occur most frequently. Increasing the multiplicity of a covalence (see p. 579) or decreasing the coordination number is accompanied by a decrease in the effective atomic radius. Some of the single-bond radii for a number of elements are quoted in Table 51 in Ångström units (10^{-8} cm.).¹²

TABLE 51. COVALENT SINGLE-BOND RADII FOR TETRAHEDRAL COORDINATION

Carbon	Nitrogen	Oxygen	Fluorine
0.77 Å.	0.74 Å.	0.74 Å.	0.72 Å.
Silicon	Phosphorus	Sulfur	Chlorine
1.17	1.10	1.04	0.99
Germanium	Arsenic	Selenium	Bromine
1.22	1.18	1.14	1.14
Tin	Antimony	Tellurium	Iodine
1.40	1.36	1.32	1.33

For the calculation of ionic radii there is no direct starting point, as is the case with covalent bonds. The earlier data were based on 1.32 Å. for the O^{--} ion and 1.33 Å. for F^- , calculated by J. A. Wasastjerna (1923) from measurements of optical refraction (p. 542). It appears, however, that these radii are somewhat too low; the values of 1.40 Å. for O^{--} and 1.36 Å. for F^- given in Table 52 are based on a division of the interionic distances in the alkali halides, and upon other considerations. From measurements of various metallic oxides and halides, the ionic radii in

TABLE 52. IONIC RADII FOR COORDINATION NUMBER SIX

Li ⁺	Be ⁺⁺			O ⁻⁻	F ⁻		
0.60 Å.	0.31 Å.			1.40 Å.	1.36 Å.		
Na ⁺	Mg ⁺⁺	Al ⁺⁺⁺	Si ⁺⁺⁺⁺	S ⁻⁻	Cl ⁻		
0.95	0.65	0.50 Å.	0.41 Å.	1.84	1.81		
K ⁺	Ca ⁺⁺	Sc ⁺⁺⁺	Ti ⁺⁺⁺⁺	Se ⁻⁻	Br ⁻		
1.33	0.99	0.81	0.68	1.98	1.95		
Rb ⁺	Sr ⁺⁺	Y ⁺⁺⁺	Zr ⁺⁺⁺⁺	Te ⁻⁻	I ⁻		
1.48	1.13	0.93	0.80	2.21	2.16		
Cs ⁺	Ba ⁺⁺	La ⁺⁺⁺	Ce ⁺⁺⁺⁺				
1.69	1.35	1.15	1.01				
		Ag ⁺ = 1.26 Å	NH ₄ ⁺ = 1.48 Å.				
Mn ⁺⁺	Fe ⁺⁺	Co ⁺⁺	Ni ⁺⁺	Zn ⁺⁺	Cd ⁺⁺	Hg ⁺⁺	Pb ⁺⁺
0.80	0.75	0.72	0.70	0.74	0.97	1.10	1.24 Å.

Table 52 have been evaluated; these are for a coordination number of six, since this is common in ionic lattices. An increase to eight is accompanied by an increase of roughly 3 per cent, while a decrease to four in the coordination number is associated with a similar decrease in the ionic radii. The values quoted in Tables 51 and 52 cannot, of course, be regarded as precise, for the additivity rule, on which they are based, is not exact; they are, nevertheless, a very good approximation in the majority of cases.¹³

It is of interest to note the influence of charge on ionic dimensions. Sodium and fluorine ions have the same electronic structures, as also have potassium and chlorine, rubidium and bromine, and cesium and iodine; the positively charged cations are, however, very much smaller than the equivalent negative ions. It is probable that in the cations the resultant positive charge causes the electronic "orbits" to be drawn inward, whereas in the anions the reverse tendency exists; the ions are thus found to be smaller and larger, respectively, than the corresponding atoms. This may be seen by comparing the radii of the positive silicon ion (0.41 Å.) and the covalent silicon atom (1.17 Å), and those of the negative chlorine ion (1.81 Å.) and the covalent atom (0.99 Å.).

It is important to note that the values recorded above really represent the radii of the atoms when connected to another atom; in other words, they are really the contributions to bond lengths. When two atoms are not united by a bond, however, the distance of closest approach observed in crystals, that is, the so-called **nonbonded radius**, is considerably increased. Every atom behaves as if it were surrounded by a sheath, between 0.5 and 1 Å. thickness, which cannot be penetrated by other atoms not joined to it by valence forces; the nonbonded distance, from center to center, of oxygen atoms, or ions, is 3.5 Å., and a similar value applies to carbon atoms. The distance between two atoms in a crystal thus forms a most valuable method of deciding whether they are joined chemically or not. In a purely electrovalent compound no ion is bound definitely to any other ion, and so the question of bonded and nonbonded radii does not arise.

Another interesting application of atomic radii is to indicate a possible means for distinguishing between covalent and electrovalent linkages. The sum of the atomic radii for covalent compounds whose constituents have the same total atomic number, that is, with the same total number of electrons, is almost constant, but when the substances are ionic the sum of the radii is not constant. In the four following substances, of total atomic number 100, the bonds are mainly covalent and the sum of the atomic radii is almost constant; the coordination number is four in each case.

AgI	CdTe	InSb	SnSn
2.81	2.80	2.79	2.79 Å.

When the bonds are ionic, however, as in the following compounds, of total atomic number 28, a definite variation is seen; the lattice coordination number for these substances is six.

NaCl	MgS	KF	CaO
2.81	2.59	2.67	2.40 Å.

Some authors have proposed to distinguish between covalent and ionic bonds by finding which set of radii gives the best agreement with experiment; the result may be misleading, however, unless the values for correct coordination number are employed.

Compounds of the AX Type.—From a consideration of the crystal structures of a large number of inorganic compounds, V. Goldschmidt (1924–29) was led to postulate the following fundamental law of crystal chemistry: *the structure of a crystal is determined by the ratio of the numbers, the ratio of the sizes, and the properties of polarization of its structural units, i.e., atoms or ions.* In addition, allowance should be made for repulsive forces in ionic compounds, especially when the cation is small, e.g., Li^+ , and the anions are in virtual contact with one another (L. Pauling, 1927). The atomic weights of the constituents, however, play no essential part in the determination of crystal structure. The value of this generalization will be illustrated by reference to simple substances of the type AX and AX_2 only, although it has been applied to compounds of greater complexity.

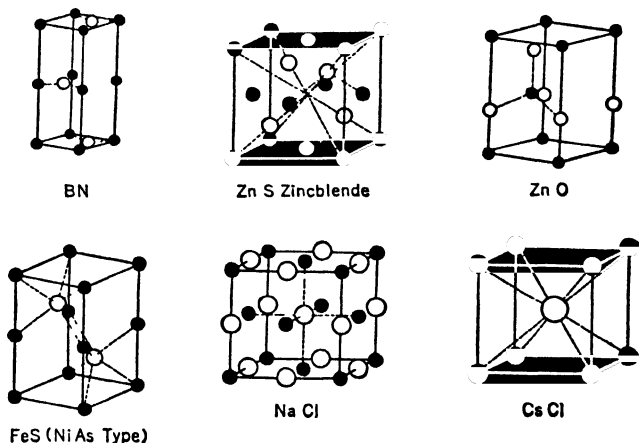


FIG. 80. Typical AX lattices

Substances of the general formula AX mostly crystallize in one of the six forms shown in Fig. 80, each lattice being named according to a familiar compound having that type of structure. Examination of the diagrams shows that the lattice coordination numbers decrease from eight in cesium chloride to three in boron nitride; the individual values are as follows:

Lattice Type	CsCl	NaCl	FeS(NiAs)	ZnO	ZnS	BN
Coordination No.	8	6	6	4	4	3

In seeking for an explanation of the difference in coordination number for various substances of the same molecular type AX , the obvious suggestion is that ionic size is the determining factor. If a small positive ion, for example, is to be surrounded by large negative ions, the coordination number must inevitably be small, but if the positive ion is relatively large and the negative ions relatively small, a large coordination number is possible. It can be readily calculated that for eight spheres, assumed to be negative ions, to surround and touch an inner one, assumed to be a positive ion, the ratio r_+/r_- must be greater than 0.73, where r_+ and r_- are the radii of the positive and negative ions, respectively; the ratios for other coordination numbers can be evaluated and the following results obtained.

Coordination No.	8	6	4	3
Ratio r_+/r_-	> 0.73	0.41–0.73	0.22–0.41	0.15–0.22

The importance of the ratio of the ionic radii in determining the crystal structure of the alkali halides may be ascertained from an examination of the data in Table 53, where the nature of the lattice type stable under normal conditions is indicated.

TABLE 53. RATIO OF IONIC RADII (r_+/r_-) FOR ALKALI HALIDES

	Lithium	Sodium	Potassium	Rubidium	Cesium
Fluoride	0.51	0.73	1.00	1.11	1.25
Chloride	0.37	0.54	0.74	0.82	0.91
Bromide	0.35	0.50	0.68	0.75	0.84
Iodide	0.31	0.45	0.61	0.68	0.75
	NaCl-type				CsCl-type

These results show that, with but few exceptions, the radius ratio appears to be the essential factor deciding the type of lattice in which the alkali halide crystallizes, and the change of rubidium chloride, bromide and iodide to a CsCl-lattice at high pressures is in harmony with their intermediate position. The abnormal behavior of some fluorides, which retain the NaCl-structure in spite of the high values of the ratio of the ionic radii, has been ascribed to the low polarizability* of the fluoride ion (see p. 540), based on the belief that high anion polarizability favors the CsCl-type of structure; this view appears to be supported by the fact that cesium iodide has the CsCl-lattice although its radius ratio is close to 0.73. According to the calculations (p. 408) of the lattice energy, however, the stability of the CsCl-lattice is to be ascribed to a high value of the van der Waals forces; in general, the magnitude of these forces increases with the size of the anion and hence the reluctance of fluorides to adopt the CsCl-structure can be understood. The retention of the NaCl-lattice in lithium chloride, bromide and iodide, in spite of the low ratio of the ionic radii, is probably due to repulsion of the anions, which are in virtual contact, so that the size of the lithium ion is effectively increased.

It may be recorded that ammonium chloride, bromide and iodide, for which the ionic ratios are the same as for the corresponding rubidium halides, normally crystallize in the CsCl-form, although at high temperatures a transition occurs

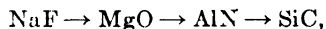
* Polarizability is a measure of the ease of deformation of the electron shells in an electric field.

to the NaCl-type of lattice. Ammonium fluoride crystallizes in the ZnO-type of lattice, but this exceptional behavior may be due to the attachment of the fluorine to nitrogen through a hydrogen bond.

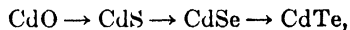
In compounds of the formula AX involving bivalent atoms the CsCl-structure appears to be unknown. The oxides, sulfides and selenides of magnesium, calcium, strontium and barium, and also the tellurides of the last three metals, crystallize in the NaCl-form. Apart from strontium and barium oxides, the ratios of the radii lie between 0.4 and 0.75, and so the lattice coordination number of six is to be expected. For the two oxides mentioned, however, the ratios are 0.86 and 1.00 respectively, and the failure to crystallize in the CsCl-form may be attributed to the small polarizability of the O^{--} ion, although other factors no doubt play a part.

The FeS- or NiAs-type of lattice is apparently restricted to compounds of the transition elements, viz., chromium, manganese, iron, cobalt, nickel, palladium and platinum, with sulfur, selenium, tellurium, arsenic, antimony and bismuth, e.g., the sulfides, selenides, tellurides and antimonides of iron, cobalt, nickel and chromium. The lattice coordination number in these substances is six, as in sodium chloride, but the NiAs-structure appears to be favored by substances in which there is a tendency for a covalent linkage to form. Since there are an insufficient number of electrons for the six coordinated atoms to be attached to the central one by ordinary electron-pair bonds, the linkages must approach metallic bonds in character. It is significant, therefore, that the compounds in this group do in fact exhibit properties, e.g., electrical conductance, opacity and lustre, which are characteristic of metals. Further, excess of one of the constituents is often able to enter the crystal lattice; such behavior is not to be expected for an ionic or simple covalent lattice.

When the ratio of the ionic radii falls below 0.4, the ZnO-structure makes its appearance as, for example, in magnesium telluride ($r_+/r_- = 0.35$), but it is also found in zinc oxide itself ($r_+/r_- = 0.63$) and in other substances for which the radius ratio is greater than 0.4. It has been mentioned that when the coordination number is four, as in the ZnO- and ZnS-types of lattice, there is a possibility of the formation of covalent linkages, but it appears that when such bonds tend to form, for example in compounds of a strongly polarizing cation, i.e., one of small size and high charge, with a highly polarizable anion, e.g., one of large size, these lattice structures are favored. The ZnO-type is probably intermediate between the ionic NaCl-lattice and the relatively covalent ZnS-structure, as may be seen from the series



where the first two, definite ionic compounds, have the NaCl-lattice, the third has a ZnO-lattice, and the fourth a purely covalent compound, has a ZnS-structure. A similar transition with increasing size, and hence polarizability, of the anion is evident in the series



where the oxide has the NaCl-lattice, the sulfide exists in both ZnO- and ZnS-types, whereas the telluride has the ZnS-structure only. As is to be expected, the ZnO- and ZnS-structures will be formed more readily in those compounds in which the electrovalence tends to pass over into a covalence; this is particularly the case for the cations Cu^+ , Ag^+ , and Hg^{++} , especially when the anion is of a strongly polarizable type, e.g., the iodide ion.

Structures of the ZnS- and ZnO-types are formed, to a large extent, between two elements, one of which is the same number of places to the left of the fourth group in the periodic table as the other is to the right, e.g., InSb, CdTe, AgI, GaAs, ZnSe and CuBr. The interatomic distances in these and similar cases (see p. 384) provide strong evidence for the existence of covalent bonds. In view of the resemblance of the structures to those of diamond the crystals are sometimes said to be *adamantine* in type.

Compounds of the AX₂ Type.—Compounds having the general formula AX₂ generally crystallize in forms based on the eight typical lattices shown in Fig. 81; the corresponding coordination numbers are given

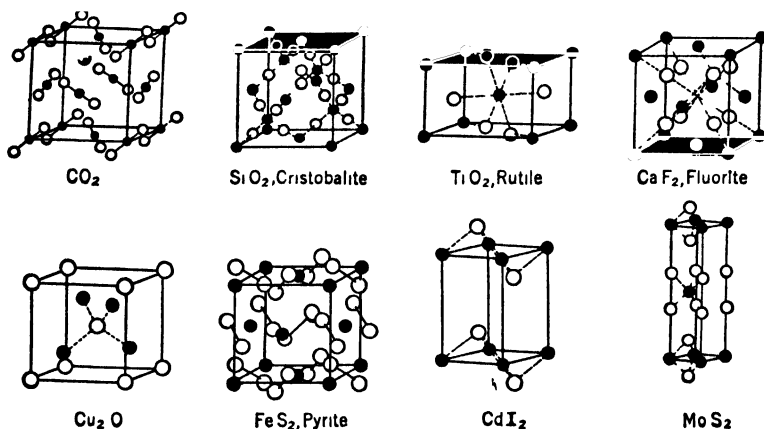


FIG. 81. Typical AX₂ lattices

below, it being noted that the number of X atoms round each atom of A must be twice as large as the A atoms about each X, because of the formula AX₂. The limiting ratios of the ionic radii r_+/r_- , based entirely on geometrical considerations, are also included. The data for the

Lattice Type	Coordination A	Number X	Ratio r_+/r_-
CaF_2	8	4	> 0.73
TiO_2 , CdI_2 , MoS_2 and FeS_2	6	3	0.41–0.73
SiO_2 and Cu_2O	4	2	0.22–0.41
CO_2	2	1	—

ratio of the ionic radii of a number of dioxides and difluorides which have the CaF_2 (fluorite) and TiO_2 (rutile) structures are given in Table 54. The fluorite structure is favored when r_+/r_- falls below 0.67, and this is close enough to the geometrical value, 0.73, to suggest that packing of ionic spheres is the essential factor in determining the lattice type. The compounds mentioned may, therefore, all be assumed to have ionic linkages. A number of sulfides and selenides of the AX₂ type, e.g., Cu_2S and Cu_2Se , where X is the metal, fall into the same classification as

TABLE 54. RATIO OF IONIC RADII FOR DIOXIDES AND DIFLUORIDES

Rutile Structures	MnO ₂	TiO ₂	OsO ₂	MoO ₂	PbO ₂	TeO ₂
	0.39	0.45	0.49	0.51	0.53	0.63
	MgF ₂	NiF ₂	CoF ₂	FeF ₂	MnF ₂	ZnF ₂
	0.56	0.56	0.59	0.60	0.62	0.62
Fluorite Structures	ZrO ₂	PrO ₂	CeO ₂	UO ₂	ThO ₂	
	0.66	0.75	0.76	0.79	0.83	
	CdF ₂	CaF ₂	HgF ₂	SrF ₂	BaF ₂	
	0.75	0.79	0.84	0.89	1.04	

the dioxides and difluorides, and so it is probable that they also have ionic lattices.

It will be observed from the tabulation on p. 388 that if ionic size were the only determining factor, the four types of structure, viz., TiO₂, CdI₂, MoS₂ and FeS₂, would be equally possible. It is apparent, however, that when ionic linkages predominate the TiO₂-structure is favored, but when there is a tendency for the electrovalence to become a covalence, e.g., when a polarizable anion is combined with a strongly polarizing cation, the molecule will have a preference for the CdI₂-lattice. The transition from the TiO₂-structure to the CdI₂-type, with a CdCl₂-lattice, which somewhat resembles that of the iodide, as intermediate, can be seen from the results in Table 55; as the polarizability of the

TABLE 55. CRYSTAL STRUCTURES OF HALIDES OF THE TYPE AX₂

	Magnesium	Cadmium	Zinc	Iron
Fluoride	TiO ₂	TiO ₂	TiO ₂	TiO ₂
Chloride	CdCl ₂	CdCl ₂	CdCl ₂	CdCl ₂
Bromide	CdI ₂	CdCl ₂ -CdI ₂ *	—	CdI ₂
Iodide	—	CdI ₂	—	CdI ₂

* Intermediate type.

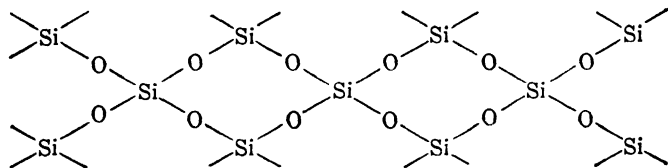
anion increases so the CdI₂-structure is favored. Other compounds of polarizable anions and polarizing cations, e.g., the disulfides, diselenides and ditellurides of quadrivalent tin, palladium, titanium, zirconium and platinum, crystallize with the CdI₂-lattice, whereas the oxides, being less polarizable, belong to the TiO₂- or CaF₂-types. An interesting case is that of calcium hydroxide which has the CdI₂-structure because of the high polarizability of the hydroxyl ion, in spite of the relatively low polarizing power of the cation.

As is to be expected, the bond distances in crystals of the CdI₂-type indicate a transition from purely ionic to partially covalent linkages, and when this tendency extends further, as a result of increased polarization, the MoS₂-lattice results. The distances between the atoms in compounds crystallizing in this form suggest that the bonds are largely covalent; examples are the sulfides, selenides and tellurides of molybdenum and tungsten. Attention may be called to the fact that the CdI₂- and MoS₂-types constitute layer lattices (p. 376).

Substances having the FeS₂-structure have many properties in common with those of the FeS(NiAs)-type already described; they consist mainly of compounds of the transitional metals with sulfur, selenium, tellurium, phosphorus, antimony and bismuth. The essential difference between the MoS₂ and the FeS₂-types is that in the former the X atoms are separate units in the

lattice whereas in the latter X_2 units only are present. The S—S distance in iron pyrites, for example, is close to that for two sulfur atoms joined by a covalent bond. The reason for the difference appears to be the valence of the metal; in the MoS_2 -group the metals are relatively large and quadrivalent and so each atom is able to associate with two separate bivalent atoms, but when the metal atom is small and bivalent it is unable to do this and, as an alternative, it joins up to the bivalent X_2 group. In view of the coordination number of six for the metal in the FeS_2 -lattice, it is probable that not all the bonds can be of the purely covalent type. These compounds, like those of the $FeS(NiAs)$ -structure, exhibit some metallic properties, and the linkages probably partake to some extent of the characteristics of those present in metals.

The transition $(TiO_2, CaF_2) \rightarrow CdCl_2 \rightarrow (CdI_2 \rightarrow (MoS_2, FeS_2))$, resulting from increasing polarization and the tendency to form covalent bonds, having been discussed, it is now necessary to consider the conditions under which the SiO_2 -structure arises. There are only two elements, viz., silicon and germanium, having atoms so small that they cannot be surrounded by six oxygen atoms; consequently, silicon and germanium dioxides crystallize with the SiO_2 -lattice in which the coordination number is four. In all the forms of silica examined, viz., α - and β -quartz, tridymite and cristobalite, the silicon atom is at the center of a regular tetrahedron with oxygen atoms at the corners, but the tetrahedra are arranged differently in the different modifications. The silicon-oxygen distance is always between 1.52 and 1.60 Å., a value which seems to suggest an ionic structure, although this is not certain; as seen above, when the coordination number is four both covalent and electrovalent bonds are possible, and it is generally believed that in silica, and in the silicates, the bonds are largely ionic in character.* An interesting feature of the structures of all forms of silica is that every oxygen atom is shared by two silicon atoms, with the result that the whole crystal is one macromolecule. The diagram below, which represents the β -cristobalite structure *projected on a plane*, will indicate how this condition arises.



When it is realized that this network is continuous in *three dimensions* throughout the crystal the physical nature of crystalline silica can be readily understood. The dextro- and levo-rotatory forms of quartz can be distinguished by their X-ray diffraction patterns; in one the SiO_4 tetrahedra which make up the crystal are connected together in the form of a right-handed, and in the other in a left-handed spiral.

* A linkage possessing double bond character, resulting from a resonance structure, has also been suggested.

Cuprous oxide and silver oxide are the only substances known to crystallize in the Cu_2O -form; this structure, with coordination number four, is certainly not dictated by the sizes of the atoms or ions, and so it must be attributed to a tendency to form covalent bonds. The interatomic distances favor this point of view.

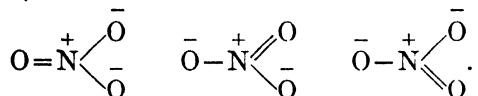
The extremity of covalent bonding among compounds of the AX_2 -type is seen in carbon dioxide, where a pair of oxygen atoms belong exclusively to one carbon atom; a true molecular lattice is formed in which the oxygen atoms are joined to carbon by purely covalent type linkages, and the unit of structure is the simple CO_2 molecule. The contrast in physical properties between carbon dioxide and its homologue silica is apparent from the fundamental difference in the structures of the crystals. In the former each molecule is held to others by relatively weak van der Waals forces, whereas in the latter the whole crystal is one large molecule. This striking change in passing from carbon to silicon may be attributed to the fact that carbon is able to form multiple bonds with oxygen, while silicon is apparently unable to do so and can only satisfy its quadrivalence in a manner similar to that depicted above for cristobalite. Nitrous oxide has an analogous electronic structure to carbon dioxide, and also has a similar crystal lattice.¹⁴

Ions of Oxyacids.—The crystal structures of many salts of the oxyacids of nonmetals, e.g., carbon, nitrogen, sulfur, phosphorus, etc., have been determined by X-rays; the anion, consisting of the central non-metallic atom surrounded by the correct number of oxygen atoms, forms a definite unit of structure, of constant size and configuration, independent of the nature of the cation. The structure of the crystal as a whole is, however, determined by the relative sizes of the anion unit and the cation. In the series of carbonates of metals of the second group, for example, magnesium carbonate has the same structure as calcite, but strontium and barium carbonates have the aragonite structure. Calcium carbonate, with a positive ion occupying an intermediate position in size, can crystallize in both forms.

The arrangement of the oxygen atoms in the anion depends on the central atom. The NO_3^- and CO_3^{--} ions are planar, the oxygen atoms occupying the corners of an equilateral triangle with nitrogen or carbon at the center. On the other hand, ClO_3^- , BrO_3^- and SO_3^{--} ions are pyramidal, the oxygen atoms being in one plane, and the other atom at the apex of the flat pyramid. In ions of the type XO_4 , e.g., PO_4^{--} , SO_4^{--} , CrO_4^{--} , ClO_4^- and MnO_4^- , the oxygen atoms are arranged tetrahedrally about the central atom; the anion structure remains unchanged when the oxygen is replaced by other elements, for example, as in BF_4^- , BeF_4^{--} and PO_3F^- . The nitrite (NO_2^-) and chlorite (ClO_2^-) ions have been found to be angular.

The interatomic distances in anions, as measured by X-rays, have proved of considerable interest. The C—O distance in the carbonate ion is 1.31 Å., and in the nitrate ion the N—O distance is 1.21 Å., whereas

the additive values for singly-bonded atoms should be 1.43 Å. and 1.36 Å., respectively. The observed distances are in fact close to those required for double bonds. Since the three oxygen atoms are symmetrically placed around the central atom and it is impossible, from valence considerations, for them all to be held simultaneously by double bonds, the actual ionic structures must resonate among three identical forms (p. 112), viz., for the nitrate ion,



All the N—O distances would be equal and would correspond more closely to the double than the single bond value. The same type of structure must hold for the CO_3^{--} anion.

Isomorphism.—The information obtained concerning the structure of anions, and Goldschmidt's generalizations relating the ratio of ionic radii and crystal type have contributed materially to the understanding of isomorphism. In order that two salts may be isomorphous it is not necessary that they should be chemically similar; the conditions of isomorphism are (i) the two compounds must have the same formula type; (ii) the respective structural units, atoms or ions, need not necessarily be of the same size in the two substances, but their *relative* sizes should be little different; (iii) the respective structural units should have the same polarization properties. The crystals will then have space lattices of the same type, with almost identical axial ratios. For the formation of solid solutions or overgrowths to occur, it is also necessary that the dimensions of the unit cells should not differ by more than about 10 per cent, otherwise entry of one substance into the space lattice of another would cause too great a distortion for stability. In this way, the necessity, long realized (p. 349), for isomorphous substances to have equal molecular volumes in order that solid solutions may be formed can be understood.

The above considerations account for many of the unexpected examples of isomorphism which have been discovered. Consider, for example, barium sulfate and potassium permanganate which have the same type of formula, XYO_4 ; the radii of the positive ions are $\text{Ba}^{++} = 1.35$ Å., compared with $\text{K}^+ = 1.33$ Å., and the sizes of the negative ions, which are both tetrahedral, are determined almost entirely by the four oxygen atoms in each case. In addition, the dimensions of the unit cells are $a = 8.90$, $b = 5.44$, $c = 7.17$ Å. for barium sulfate, compared with 9.10, 5.69 and 7.40 Å., respectively, for potassium permanganate. The latter salt is, therefore, able to form overgrowths on the former. Other examples of unusual types of isomorphism, e.g., KClO_4 and KBF_4 ; KIO_3 and CaTiO_3 ; K_2SO_4 , K_2BeO_4 and $\text{K}_2\text{PO}_3\text{F}$, can be accounted for in a similar manner.

Sodium nitrate and calcium carbonate, the latter as calcite, are isomorphous, but potassium nitrate is isomorphous with aragonite; the increase in the size of the cation, and hence in its ratio to that of the anion, causes the change of crystal form. It is this difference in ionic dimensions which explains why so-

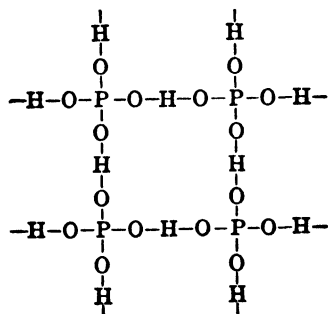
dium and potassium salts are not all isomorphous with one another; the same argument applies to calcium and magnesium salts. The compounds K_2PtCl_6 and $\text{K}_2\text{Pt}(\text{SCN})_6$ belong to different crystal systems, although they undoubtedly have similar chemical constitutions; the size and polarizability of the coordinated groups are probably the determining factors. The different crystal forms of K_2SnCl_6 and $\text{K}_2\text{Sn}(\text{OH})_6$ may also be explained by these factors.

Sodium chlorate is not isomorphous with either nitrate or carbonate, and it has an entirely different type of structure; the ClO_3^- ion, as noted above, is pyramidal whereas the others are planar. The necessity for ions to have similar stereochemical form introduces another condition for isomorphism, in addition to the three already mentioned.

Lithium and sodium chlorides form a continuous series of solid solutions when crystallizing from the molten mixture, and so also do the chlorides of sodium and potassium, all of which have the NaCl -type of lattice; nevertheless, solid lithium and potassium chlorides are not miscible with each other. The size of the unit cell is 5.14 Å. for lithium, 5.63 Å. for sodium, and 6.28 Å. for potassium chloride; the difference between the first and second, and between the second and third is approximately 10 per cent, but the third is about 20 per cent greater than the first, and solid solution of the chlorides cannot result. Analogous behaviour has been observed with other alkali halides even though they have the same crystal lattice.

An exceptional type of isomorphism is that between CaF_2 and YF_3 , and between PbF_2 and BiF_3 ; these substances all have the fluorite-type of lattice, and in the XF_3 compounds the additional fluorine atom is introduced at the center and at the middle of the sides of the unit cell (see Fig. 81) without any appreciable change in its dimensions.

Hydrogen Bonds.—In the salts of certain anions containing hydrogen, e.g., H_2PO_4^- and HCO_3^- , two oxygen atoms from neighboring groups approach more closely, viz., 2.5 to 2.8 Å., than is possible in the absence of bonding (p. 384); it is probable, in such cases, that the two oxygen atoms are linked by a hydrogen bond (p. 113). The positions of the hydrogen atoms cannot be determined by X-ray diffraction, and it is inferred that each lies between a pair of oxygen atoms. The arrangement of the phosphorus and oxygen atoms in the dihydrogen phosphate ion, H_2PO_4^- , has been determined and a unit of structure, involving hydrogen bonds, has been proposed; the following diagram represents a projection on to the plane of the paper.

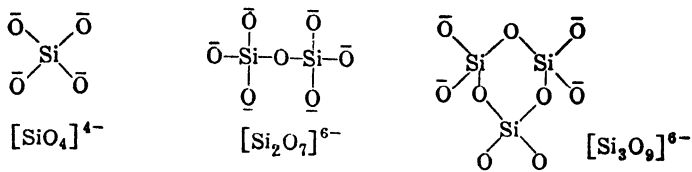


Since every hydrogen is shared between two oxygen atoms, and there are four of the latter surrounding each atom of phosphorus, the net formula is H_2PO_4^- . A similar structure, involving $-\text{O}-\text{H}-\text{O}-$ chains, has been suggested for the bicarbonate ion in crystals, and the same grouping probably occurs in other compounds, both inorganic and organic, e.g., boric acid, H_3BO_3 ; telluric acid, H_6TeO_6 ; the mineral *diaspore*, AlHO_2 ; resorcinol, $m\text{-C}_6\text{H}_4(\text{OH})_2$; and oxalic acid, $\text{H}_2\text{C}_2\text{O}_4$.

The X-ray studies of ice have shown that each oxygen atom is surrounded by four other oxygen atoms arranged tetrahedrally, the distance between the centers of two adjacent atoms being 2.76 Å. The pairs of atoms are evidently linked by hydrogen bonds, as stated on page 115. The question arises in connection with the structure of water, and of similar substances, as to whether the hydrogen atom is to be found midway between the two oxygen atoms or whether it is closer to one of them. There is some evidence that in ice the hydrogen atom is actually about 1.00 Å. from one oxygen atom and 1.76 Å. from the other. The vibration frequency of the O—H bond, as determined from molecular spectra (see Chapter VIII), is very little different in ice and water vapor. In the latter the O—H bond length is 0.96 Å., and the frequency change indicates an extension to about 1.00 Å. It appears, therefore, that in ice each oxygen atom is associated closely with two hydrogen atoms, and less intimately with two others which are closely attached to other oxygen atoms. In a sense, therefore, each water molecule in ice is a discrete entity, although it is attached to other molecules in such a way as to make the whole crystal virtually one large molecule.

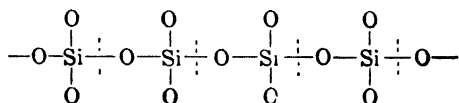
The Structure of Silicates.—A highly significant contribution to the chemistry of the silicates has been made as a consequence of their X-ray investigation, chiefly by W. L. Bragg (1926 *et seq.*). One result of the work has been the simple interpretation of the formulae of many of the naturally occurring mineral silicates. The ratio of silicon to oxygen in these compounds has a number of different values, from 1:4 in the simple orthosilicates, to 1:2 in silica itself, so that all silicates may be regarded as falling between these two extremes. It was seen (p. 390) that in silica each silicon atom (or ion) is surrounded tetrahedrally by four oxygen atoms (or ions); this arrangement persists throughout all the silicates, the complexity arising because of the possibility, as in silica, of an oxygen atom being shared by two silicon atoms. As this sharing can occur in different ways, silicate ions of varying complexity are obtained. The cations are generally distributed symmetrically throughout the crystal so as to provide greatest stability.

The simplest silicate ion is obviously $[\text{SiO}_4]^{4-}$, which gives rise to the orthosilicates, e.g., Mg_2SiO_4 , Be_2SiO_4 , $(\text{AlF})_2\text{SiO}_4$, $(\text{Al}_2\text{O})\text{SiO}_4$. The sharing of one oxygen atom by two silicon atoms, while retaining the tetrahedral configuration of each, gives the ion $[\text{Si}_2\text{O}_7]^{6-}$, as in $\text{Ca}_2\text{MgSi}_2\text{O}_7$, $\text{Zn}_4(\text{OH})_2\text{Si}_2\text{O}_7 \cdot \text{H}_2\text{O}$, whereas when three oxygens are shared by two silicon atoms to form a ring structure the ion $[\text{Si}_3\text{O}_6]^{6-}$, as in $\text{BaTiSi}_3\text{O}_9$, results. These ions may be represented diagrammatically by the following planar projections.

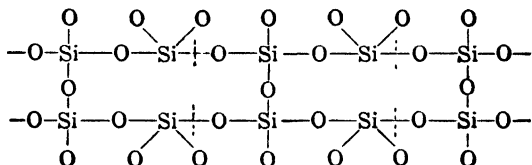


A ring of six tetrahedra would give the ion $[\text{Si}_6\text{O}_{18}]^{12-}$ only found in *beryl*, $\text{Be}_3\text{Al}_2\text{Si}_6\text{O}_{18}$.

The next stage of complexity arises when a number of tetrahedra are joined into a long chain, by an oxygen atom held in common by each adjacent pair; thus,



giving the general formula $[(\text{SiO}_3)_n]^{2n-}$. The formation of a double chain



leads to ions of the type $[(\text{Si}_2\text{O}_7)_n]^{6n-}$. These two structures are found in the important minerals known as *pyroxenes* and *amphiboles*, respectively. The long chains or bands of silicate ions are held side by side in the crystal by the cations which interleave them, but as the forces are relatively weak, compared with those existing in the chains, the substances are of a fibrous nature.

The lateral development of the chains leads to the formation of a sheet of linked tetrahedra, giving an ion of the formula $[(\text{Si}_2\text{O}_5)_n]^{2n-}$; the layers are held together by cations, and as might be expected the crystals cleave readily into flakes. *Talc*, $\text{Mg}_3(\text{OH})_2\text{Si}_4\text{O}_{10}$, is an example of this group of silicates. The micas also belong to the same category, although in these substances one-quarter of the silicon atoms are replaced by aluminum, as in *muscovite*, $\text{KAl}_2(\text{OH})_2(\text{Si}_3\text{AlO}_{10})$. Other cases of replacement of silicon by aluminum occur among the silicates, but the introduction of each aluminum atom results in an increase of the negative charge by unity, since the latter requires one more electron than a silicon atom to complete its electron octet. It is therefore necessary to introduce an additional positive charge into the molecule to ensure neutrality, as is evident from the formulae of talc and muscovite. The net charge of the cationic portion is +5 to balance the -5 of the $(\text{Si}_3\text{AlO}_{10})$ group.

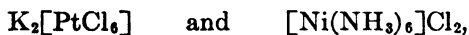
It will be noted that in several of the examples quoted the OH group appears in the cation portion. It has been found by X-ray examination that not only can the chains, bands or sheets of silicate be held together by simple cations, but the positively charged units are often groups such as $(\text{AlO})^+$, $(\text{AlF})^{++}$ or $(\text{AlOH})^{++}$. In addition to these, layers of $\text{Mg}(\text{OH})_2$ or MgF_2 frequently occur, e.g., in the group of minerals represented by the general formula $\text{Mg}(\text{F}, \text{OH})_2 \cdot n\text{MgSiO}_4$.

The extension of the silicon-oxygen tetrahedra in three dimensions leads, of course, to structures similar to those of the forms of silica. It is possible,

however, for some of the silicon atoms to be replaced by aluminum; this causes the three-dimensional network to become anionic, and so it is able to combine with cations, just as do the various silicate ions. One univalent positive ion must be present for each aluminum atom, and the sum of the silicon and aluminum atoms must be equal to half the number of oxygen atoms, so as to preserve the silica-like structure. Many minerals such as the feldspars and zeolites, consist of aluminosilicates of this type, e.g., the feldspar *orthoclase*, KAlSi_3O_8 , and the zeolite *natrolite*, $\text{Na}_2\text{Al}_2\text{Si}_2\text{O}_{10} \cdot 2\text{H}_2\text{O}$. In the zeolites the SiO_4 and AlO_4 tetrahedra form an open network, the cations and water molecules being accommodated in the relatively large spaces thus formed. This probably accounts for the ease with which the cations can be exchanged with other positive ions in solution, and with which the water molecules can be replaced by other molecules (p. 1204). In the mineral *sodalite*, $\text{Na}_4\text{Al}_3\text{Si}_3\text{O}_{12}\text{Cl}$, an additional sodium ion is present to compensate for the chloride ion which occupies one of the holes in the cage-like network of tetrahedra. It is of interest to note that the structure of *ultramarine* is similar to that of *sodalite*, except that complex sulfur ions, S_x^- , take the place of the chloride ions.¹⁵

The complex borates have not yet been subjected to thorough examination by X-ray methods, but it appears reasonably certain that they resemble the silicates in forming chains and nets. The lattice coordination number of the boron atom is three, and the three oxygen atoms surrounding it lie in one plane (W. H. Zachariasen, 1931). Studies have been made of the heteropoly acids and their salts, e.g., phospho- and silico-tungstic acids, particularly by J. F. Keggin (1934), and the results have thrown light on the structural chemistry of these complex substances. A number of empirical rules put forward by L. Pauling (1929) have been employed in order to indicate preliminary structures of crystals containing highly complex ions of the type under consideration.¹⁶

Coordination Compounds.—The work on coordination compounds has fully confirmed the fundamental basis of Werner's theory concerning the structure of such compounds. A striking fact is the similarity in crystal structure of



both of which have the fluorite lattice (p. 388); analogous substances, e.g., $\text{Co}(\text{NH}_3)_6\text{I}_2$, $(\text{NH}_4)_2\text{SiF}_6$ and $(\text{NH}_4)_2\text{SnCl}_6$, have lattices of the same type. In each case the bivalent group MR_6 , whether positive or negative, constitutes a unit of crystal structure; the six R groups are arranged in a regular octahedron about the central atom M, in agreement with Werner's original views. Although $\text{K}_2\text{Pt}(\text{SCN})_6$, $\text{K}_2\text{Sn}(\text{OH})_6$, $(\text{NH}_3 \cdot \text{C}_2\text{H}_5)_2\text{SnCl}_6$ and $\text{Ni}(\text{NH}_3)_6(\text{NO}_3)_2$ do not crystallize in the fluorite form, they retain the octahedral configuration of the complex ion.

The planar structure proposed by Werner for the 4-covalent platinum compounds has been confirmed, e.g., K_2PtCl_4 (R. G. Dickinson, 1922) and $\text{Pt}(\text{NH}_3)_4\text{Cl}_2 \cdot \text{H}_2\text{O}$ (E. G. Cox, 1932); the analogous palladium compounds are structurally similar. The question of whether compounds of this type, with a Werner coordination number of four, are always planar, or whether they are sometimes tetrahedral, has attracted attention in

recent years, and the X-ray method has been employed in a variety of instances to supply an answer. In a number of derivatives of bivalent nickel, palladium, platinum, copper and silver, the coordinated groups are undoubtedly arranged in a planar manner round the central atom, but in trimethyl platinic chloride, $\text{Pt}(\text{CH}_3)_3\text{Cl}$, in which the platinum is quadrivalent, it appears that a planar configuration is improbable. Further, in the cuprocyanide ion, $\text{Cu}(\text{CN})_4^{--}$, the valence distribution is tetrahedral about the central *univalent* copper atom, just as it is in the complex cyanides $\text{M}(\text{CN})_4^{--}$ where M is Zn, Cd or Hg. The quadrivalent compounds of tin and lead have a tetrahedral configuration, as might be expected by analogy with carbon, but the SnCl_4^{--} ion, in which tin is bivalent, is planar. There is clearly a connection between the principal valence of an atom and the arrangement of its chemically coordinated groups.

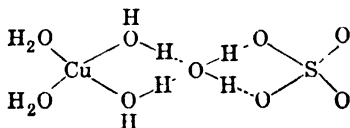
Interesting results have been obtained with complex salts in which the central atom has apparently an odd coordination number, but in most cases the correct value is even, the odd number resulting from fortuitous circumstances, e.g., packing in the crystal.* In the ions $\text{Tl}_2\text{Cl}_9^{--}$ and $\text{W}_2\text{Cl}_9^{--}$ each of the metal atoms is surrounded octahedrally by six chlorine atoms, but the two octahedra have a face in common; that is to say, three chlorine atoms are attached to both metal atoms. In this manner nine chlorine atoms are joined to two thallium or tungsten atoms each of which has a coordination number of six. In Cs_3CoCl_5 the complex ion is actually $[\text{CoCl}_4]^{--}$; the fifth chlorine is quite differently placed from the other four, and the crystal structure thus involves three cesium ions, one CoCl_4^{--} ion and one chloride ion (H. M. Powell, 1935-36). It is of interest to record in connection with the subject of complex ions that $\text{Cs}_3\text{As}_2\text{Cl}_9$ has been proved to consist of molecular AsCl_3 units together with cesium and chloride ions, viz., $(\text{Cs}^+\text{Cl}^-)_3(\text{AsCl}_3)_2$, and the lattice of $\text{NH}_4\text{Pb}_2\text{Br}_5$, and the corresponding potassium and rubidium salts, are made up of ammonium, or alkali metal, and bromine ions and molecular covalent lead bromide, that is $\text{NH}_4^+\text{Br}^-(\text{PbBr}_2)_2$. The salt $\text{K}_2\text{CuCl}_4 \cdot 2\text{H}_2\text{O}$, similarly, has the constitution $(\text{K}^+\text{Cl}^-)_2\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$, the $\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$ portion having the same covalent structure as the simple salt. Substances of this type, formed between definite stoichiometric amounts of two molecular species, which owe their stability to packing in the crystal lattice, and not to ordinary valence forces, are called **lattice compounds**. A number of "molecular compounds," e.g., $\text{CHI}_3 \cdot 3\text{S}_8$, the hydrates of the inert gases, and the alums, probably belong to this category, for they exist only in the solid state.¹⁷

Water of Crystallization.—Molecules of water can form part of the structure of crystals in various ways. In some compounds they are apparently joined to the central atom by covalent, or possibly ionic (see p. 622), bonds; in this respect molecules of water may partly or

* $(\text{NH}_4)_3\text{ZrF}_7$ and $(\text{NH}_4)_2\text{TaF}_7$, and analogous substances, are probably true odd coordination compounds.

wholly replace those of ammonia in metal-ammines. Substances for which this behavior has been confirmed by X-ray investigation are $\text{Co}[(\text{NH}_3)_5(\text{H}_2\text{O})]\text{X}_3$ and $\text{Co}[(\text{NH}_3)_4(\text{H}_2\text{O})_2]\text{X}_3$. In the extreme case, when all the NH_3 groups have been replaced by H_2O , compounds such as $\text{Zn}(\text{BrO}_3)_2 \cdot 6\text{H}_2\text{O}$ and $\text{Ni}(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$ are obtained in which the octahedral arrangement of the coordinated groups is still maintained; the hydrated salts should thus be represented as $[\text{Zn}(\text{H}_2\text{O})_6](\text{BrO}_3)_2$ and $[\text{Ni}(\text{H}_2\text{O})_6](\text{NO}_3)_2$, respectively. X-ray measurements on $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$ and $\text{AlCl}_3 \cdot 6\text{H}_2\text{O}$ indicate that these hydrated salts are constructed in a similar manner. In the group of compounds $\text{M}^{\text{II}}\text{M}^{\text{IV}}\text{X}_6 \cdot 6\text{H}_2\text{O}$, where M^{II} is Mg, Mn, Zn, Ni, etc., and M^{IV} is Si, Sn, Pd, Pt, etc., and X is a halogen, the structure is generally of the CsCl-type, with $[\text{M}^{\text{II}}(\text{H}_2\text{O})_6]^{++}$ as one unit and $[\text{M}^{\text{IV}}\text{X}_6]^{--}$ as the other. The molecules of water are arranged octahedrally about the M^{II} ion and are presumably attached to it by valence bonds.

A different type of structure appears in $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ and $\text{NiSO}_4 \cdot 7\text{H}_2\text{O}$, for some of the molecules of water surrounding the central metal ion are joined by hydrogen bonds to others which are connected with the sulfate ion. In $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, for example, four molecules of water are arranged in a plane about the copper ion, to which they are apparently connected by valence bonds, while in $\text{NiSO}_4 \cdot 7\text{H}_2\text{O}$ six of the water molecules are disposed octahedrally around the nickel ion. The additional water molecule in each case is connected tetrahedrally to two water molecules of the first type and to two oxygen atoms of the sulfate ion by hydrogen bonds. The structure of $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ is thus shown schematically by the following planar representation (C. A. Beevers, 1934-35).



The alums are somewhat similar to these sulfates as regards the disposition of the water molecules. In $\text{M}^{\text{I}}\text{M}^{\text{III}}(\text{SO}_4)_2 \cdot 12\text{H}_2\text{O}$, it appears that six molecules of water are definitely associated with the M^{III} ion, in an octahedral configuration, while each of the other six molecules acts as a link between the M^{I} ion, the SO_4^{--} ion, and the water molecules of the former group.

In oxalic acid dihydrate, $\text{H}_2\text{C}_2\text{O}_4 \cdot 2\text{H}_2\text{O}$, the molecules of water of crystallization are apparently joined to the oxygen atoms of oxalic acid by hydrogen bonds; this mode of attachment thus appears to be relatively common in acids and salts containing oxygen.¹⁸

It is not always safe to argue by analogy concerning the structures of salt hydrates. The salt $\text{KAuBr}_4 \cdot 2\text{H}_2\text{O}$, for example, contains the planar AuBr_4^- ion, but the two molecules of water do not appear to be attached either to this or to the potassium ion (E. G. Cox, 1936); they merely fit into vacant spaces in the crystal lattice. An extreme type of this be-

havior is shown by many, although not all, zeolites; the X-ray evidence suggest that molecules of water, singly or in chains, are accommodated in holes or channels in a firm structure of SiO_4 and AlO_4 tetrahedra. The water can be removed from such substances without destruction of the crystal form.*¹⁹

ORGANIC COMPOUNDS

Organic Compounds: Aliphatic.—Relatively few complete structure determinations of aliphatic compounds have been made, but interesting data of a more qualitative nature are available concerning long chain compounds. Since single crystals are generally not readily prepared, the X-ray diffraction measurements have been made by smearing the substance, or allowing a drop of liquid to solidify, on a glass plate; in this manner a powder-type of photograph is obtained. The earliest observations of this kind were made by L. de Broglie and E. Friedel (1923), but almost simultaneously a more systematic and extended investigation was reported by S. H. Piper and E. N. Grindley (1923), and by A. Müller and G. Shearer (1923). From the lines in the X-ray photographs the dimensions, d_1 , d_2 and d_3 , of the unit cell were calculated; the results for some fatty acids with an even number of carbon atoms are recorded in Table 56. The d_1 spacing increases steadily as the length of the hydro-

TABLE 56. MAIN SPACINGS IN CRYSTALS OF FATTY ACIDS

Acid	Number of Carbon Atoms	d_1	d_2	d_3
Capric	10	23.2 Å.	— Å.	— Å.
Lauric	12	27.0	4.11	3.68
Myristic	14	32.2	4.12	3.72
Palmitic	16	34.7	4.08	3.65
Stearic	18	38.7	4.05	3.62
Behenic	22	47.8	4.10	3.66

carbon chain increases, but the d_2 and d_3 values remain almost constant. This result suggests that the fatty acids consist of long chains which pack side by side in the crystal, and the whole structure is made up of a number of layers of this kind placed one on the other. The chains are oriented perpendicular, or nearly perpendicular, to the plane of the sheets, the distance between which gives the d_1 spacing; it follows, therefore, that increase of chain length will produce a regular increase in this quantity. The d_2 and d_3 spacings, on the other hand, are presumably determined by the thickness of the molecules, and these should be independent of the number of carbon atoms in the chain; the area of cross section of a single molecule is thus about 16 Å^2 . The relation of d_1 to the length of the chain will be considered shortly.

Similar measurements were made subsequently with other long chain compounds. For a series of saturated hydrocarbons, for example, A.

* The term **zeolitic water** is used to describe molecules of water in any system which is **bivariant**, instead of univariant, on dehydration (cf. p. 781); the green hydrate $\text{Cr}_2(\text{SO}_4)_3 \cdot 15\text{H}_2\text{O}$ is quoted as an illustration.

Müller and W. B. Saville (1925) obtained the d_1 spacings given in Table 57; the d_2 and d_3 values were similar, as might be expected, to those for the

TABLE 57. LATTICE SPACING FOR LONG CHAIN PARAFFINS

Formula	d_1	Formula	d_1
$C_{17}H_{36}$	24.3 A.	$C_{23}H_{48}$	32.2 A.
$C_{18}H_{38}$	25.9	$C_{24}H_{50}$	33.0
$C_{19}H_{40}$	26.9	$C_{27}H_{56}$	37.1
$C_{20}H_{42}$	28.0	$C_{31}H_{64}$	43.0
$C_{21}H_{44}$	29.4	$C_{35}H_{72}$	47.7

fatty acids. If the d_1 spacing for a hydrocarbon is compared with that for a fatty acid with the same number of carbon atoms, it is seen that the latter is almost twice as great as the former. Further, the average increase of d_1 for each additional carbon atom in the chain is approximately 2.05 Å. in the fatty acids, and 1.3 Å. in the hydrocarbon series.

The difference in the d_1 spacing, which corresponds to the chain length, is undoubtedly connected with the fact that fatty acids tend to form double molecules, stable not only in the solid but also in the liquid and vapor phases (cf. p. 319), by the sharing of hydrogen bonds between the carboxyl groups. Assuming for the present that the hydrocarbon chain can be represented by a straight line, and that these chains are perpendicular to the crystal plane, the arrangement of the molecules in fatty acids and in hydrocarbons would be as shown in Fig. 82; the $-\text{COOH}$ groups

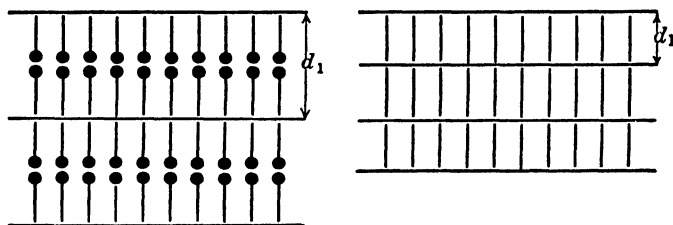


FIG. 82. Molecular arrangement in fatty acids (left) and hydrocarbons (right)

are represented by black circles. The X-ray spectra of long chain compounds are evidently determined by the planes indicated by horizontal lines; if this is the case it can be shown that the presence of powerful scattering centers midway between them, such as would occur where two $-\text{COOH}$ groups were situated, should result in the odd order spectra being weak. The actual intensities of the lines show this alternation, and so the presence of the extended double molecule may be regarded as confirmed. If the carboxyl group is esterified the bimolecular association no longer occurs, and the d_1 spacing should be of the same order as for the hydrocarbons; this is almost exactly true for the ethyl esters. Dibasic acids probably form associated double molecules, but as they have a carboxyl group at each end, the association will extend through the whole crystal; the pattern, however, repeats itself with every single molecule

(Fig. 83), and d_1 should be determined by a single, and not a double, hydrocarbon chain. The measurements on dicarboxylic acids provide complete support for this argument.

Since it may be regarded as established that the fatty acids are arranged in bimolecular layers, the increase in the d_1 spacing for each additional carbon atom is $2.05/2 = 1.02$ Å., which is still different from the value 1.3 Å. observed with the paraffin hydrocarbons. A "straight" hydrocarbon chain is most probably a zigzag of carbon atoms arranged at angles of $109^\circ 28'$, the tetrahedral angle, to each other, as shown in Fig. 84, in which the atoms are assumed to be in one plane. If the diameter of the carbon atom is 1.54 Å., as in diamond, the increase in chain length resulting from the addition of two CH_2 groups to the chain is seen to be 2.52 Å., or 1.26 Å. per carbon atom. The value is very close to that found for the paraffins, and also for ethyl esters of fatty acids, acid potassium salts, ketones and primary alcohols. It appears, therefore, that in all these substances the hydrocarbon chain is a zigzag of carbon atoms, or more correctly CH_2 groups, perpendicular to the X-ray reflecting planes of the crystal.

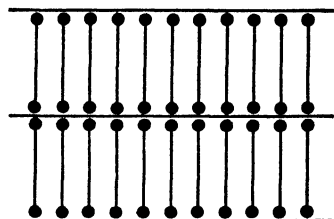


FIG. 83. Molecular arrangement in dibasic acids

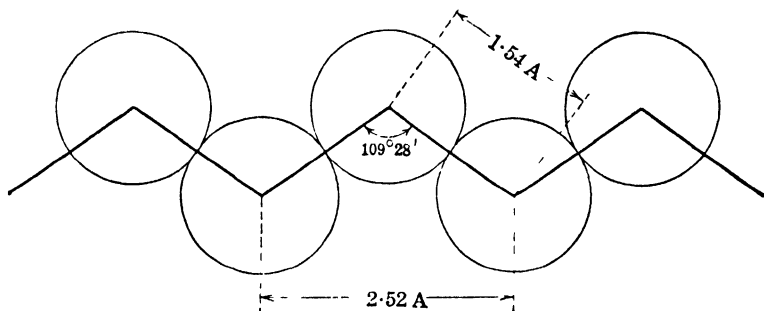


FIG. 84. Zigzag chain of carbon atoms

By using a single crystal of the hydrocarbon $\text{C}_{29}\text{H}_{60}$, A. Müller (1928) obtained an X-ray rotation spectrum showing reflections from atom-bearing planes; in this way the positions of the atoms were identified and shown to form a zigzag arrangement as suggested above. The distance between two corresponding carbon atoms was found to be 2.54 Å., equivalent to 1.27 Å. increase in length per carbon atom, in excellent agreement with the calculated value. Similar results were also recorded by J. Hengstenberg (1928) for a single crystal of $\text{C}_{36}\text{H}_{72}$.

Tilted Chains.—The explanation offered for the apparently small increment for each additional CH_2 group in the fatty acids, also found in dibasic acids, iodides and nitriles, is that the molecular chains are not perpendicular

to the reflecting planes of the crystal, but are tilted through an angle of about 30° . This suggestion is in harmony with measurements made on single crystals of some fatty acids (A. Müller, 1927). The addition of a carbon atom increases the chain length by 1.27 \AA , but the d_1 spacing will be increased by a smaller amount, viz., $d_1 \cos \alpha$, where α is the angle of tilt. It has been shown by T. Malkin (1931), following earlier suggestions by A. Müller (1929) and others, that the assumption of tilted chains provides an interpretation of the well-known alternation in the physical properties of many long-chain compounds; it is significant that marked alternation occurs chiefly in series of compounds having the shortened d_1 spacings, but it is small for paraffins, methyl ketones, primary alcohols and ethyl esters. It seems, therefore, that appreciable alternation in physical properties and tilt of the hydrocarbon chain may be related; this can be shown with the aid of Fig. 85 which involves a reasonable assumption

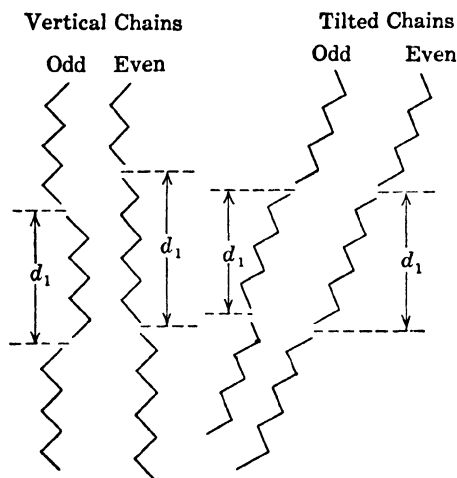


FIG. 85. Vertical and tilted hydrocarbon chains

concerning the orientation of the ends of the chains with respect to each other. It will be seen that in the vertical chains there is no difference in the density of packing in the terminal planes, i.e., in the spacings about the dashed lines. For the tilted chains, however, the packing densities in the terminal planes are obviously different. The occurrence of alternate wide and narrow terminal planes in the tilted chains with an odd number of carbon atoms, whereas in the even series they are all of the same width, should lead to an alternation in the d_1 spacings as the hydrocarbon chain is lengthened. As a general rule, as required by the theory, the d_1 spacings in the chains with odd and even numbers of carbon atoms fall into two separate series; this is clear from the following data for the fatty acids (F. Francis, S. H. Piper and T. Malkin, 1930).

	C_{12}	C_{14}	C_{16}	C_{18}	C_{20}	C_{22}
Even Series	27.4	31.5	35.6	39.9	44.2	48.5 A.
	C_{13}	C_{15}	C_{17}	C_{19}	C_{21}	
Odd Series	30.4	34.2	38.6	43.1	47.8 A.	

It must be emphasized, in conclusion, that the foregoing results refer to measurements made with the stable forms of the various substances; many long chain compounds exist in more than one crystalline form, and unless the analogous stable forms are compared misleading conclusions concerning alternation may be reached.²⁰

Aromatic Compounds.—The work in this field has been carried out with single crystals, so that detailed analysis of the structures has been possible; the results have, in general, confirmed the views of organic chemists, although some novel information has been obtained. X-ray diffraction measurements with hexamethylbenzene crystals established clearly the planar, regular hexagonal structure of the benzene ring (Mrs. K. Lonsdale, 1929); this was later confirmed with hexachlorobenzene, and since 1933 J. M. Robertson and others have applied the method of Fourier synthesis (p. 362) to determine the detailed structures of the following aromatic hydrocarbons: naphthalene, anthracene, durene, dibenzyl, chrysene, diphenyl, *p*-diphenylbenzene, 1:3:5-triphenylbenzene, stilbene, tolan, etc. The biphenyl molecule is planar, in harmony with modern views, and so also are stilbene, which has the *trans*-configuration in the crystal, and tolan, which is linear because of the triple bond. In bibenzyl, on the other hand, the two benzene rings are in different, although parallel, planes. All benzene and similar rings are regular hexagons, and the distance between two adjacent carbon atoms in the ring is always 1.40 ± 0.01 Å., compared with 1.42 Å. in graphite. The data support the views concerning the resonance structure of the benzene nucleus (p. 112). The ring in benzoquinone is not a regular hexagon; in this molecule resonance, leading to the equivalence of the six carbon-carbon bonds, cannot occur and so a "quinonoid" structure, involving two double and four single bonds in the ring, must result. The distances obtained by applying the Fourier method are 1.50 Å. for the four C—C bonds adjacent to the $>C=O$ groups, and 1.32 Å. for the two C=C linkages; these are very close to the values generally obtained for single and double carbon bonds, respectively.²¹

Fiber Structure.—Information of exceptional interest has followed from the application of knowledge gained in the field of organic chemistry to the interpretation of X-ray patterns given by fibrous materials, such as cellulose, proteins, rubber and artificial long chain polymers. When taking a rotation photograph (p. 372), a single crystal is mounted with an axis perpendicular to the X-ray beam and rotated about this axis so as to present different aspects to the beam; if a number of small crystals, all arranged with a given axis in the same direction but otherwise having random orientation, form a stationary bundle, the result on exposure to X-rays should correspond to a rotation photograph. When a fine beam of monochromatic X-rays is passed through a bundle of parallel fibers of

ramie, silk or hair, in a direction perpendicular to the length of the fiber, a pattern is in fact obtained having the essential characteristics of an

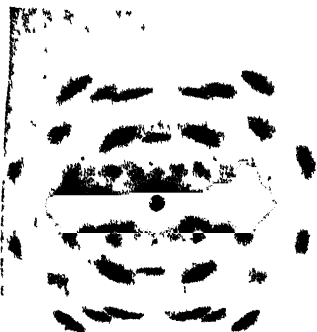


Fig 86. X-ray diffraction pattern of ramie fiber

ideal rotation photograph given by a single crystal (Fig. 86). It follows, therefore, that fibers of this nature are not amorphous, but consist, to a great extent, of a number of pseudo-crystals or **crystallites** arranged so that one and the same crystallographic axis is always approximately parallel to the length of the fiber. Perfect orientation is, of course, hardly to be expected, and this is apparent from the diffuse nature of the "spots" in the photograph; the more random the arrangement the more does the pattern broaden out into a series of diffuse rings similar to a powder photograph. The production of faint rings by nearly all fibers studied suggests that some unoriented material is always present.

The X-ray reflections on or near the "equator" of fiber photographs are always broader and more diffuse than those at or near the "meridian," and so it can be shown, as indicated below, that the dimensions of the crystallites along the direction of the fiber must be much greater than at right angles. It is this fact which accounts for the essential "fibrous" nature of the substances. If an X-ray pattern is taken with the beam parallel to the fiber length, only a series of rings are obtained; the arrangement of the crystallites must, therefore, be completely random apart from their alignment in the direction of their length.

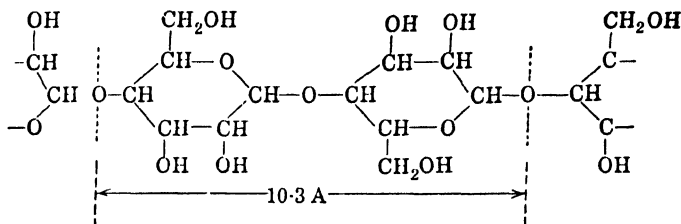
It is recorded that the X-ray diffraction by fibers was demonstrated by S. Nikishawa and S. Ono in 1913, and by P. Scherrer in 1918, although the first attempt to interpret the results was made by R. O. Herzog, W. Jancke and M. Polanyi (1920-22); important work was also carried out by O. L. Sponsler (1922-26), but the modern development of the field of fiber structure is to be attributed to H. Mark and K. H. Meyer (1928) who applied the information gained concerning the chemistry of cellulose to interpret the X-ray data. As a result it has been concluded that fibrous materials consist of bundles made up of definite long chain chemical units; the direction of the length of the chain is that of the crystallites in the fiber. Much important progress has been made in the study of fibers, and a brief survey of some of the more interesting conclusions will now be given.

Cellulose.—From the position of the spots in the diffraction pattern of cellulose, M. Polanyi (1921) estimated the size of the "unit cell," and later Meyer and Mark showed that it was monoclinic and gave its dimensions as

$$a = 8.3 \text{ \AA.} \quad b = 10.3 \text{ \AA.} \quad c = 7.9 \text{ \AA.} \quad \beta = 84^\circ,$$

the b axis being parallel to the length of the fiber. The formula of cellulose is

known to be $(C_6H_{10}O_5)_n$, and chemical examination has shown it to consist of a number of glucose residues joined together in the 1:4-positions (W. N. Ha-worth, 1926); the X-ray data suggest that the cellulose fiber is made up of a number of chains of such residues. From the known atomic dimensions (p. 383) the distance 10.3 Å. in the direction of the fiber axis is equivalent to two glucose residues, and so the pattern must repeat itself for each two such residues; the structural unit in the cellulose chain can, therefore, be represented by the formula between the dotted lines.



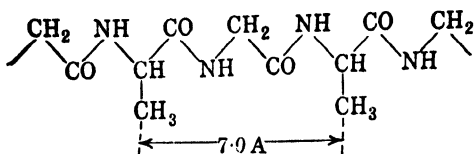
There are actually four such chains, parallel to each other, involved in the unit cell, thus accounting for the a and c dimensions, but as some of the chains are shared between several adjacent cells there are four glucose residues on the average in each. The detailed arrangement in the cell need not be discussed here, but sufficient has been said to show that a cellulose fiber consists of a bundle of long chains of double glucose units; the direction of the chain corresponds with the length of the crystallites and of the fiber. Other fibers studied have been found to be built up in an analogous manner.

The same essential structure has been obtained for celluloses of different origin, thus showing that, chemically at least, they are identical. There is, however, some difference in the arrangement of the crystallites; ramie, for example, gives a very clear X-ray diffraction pattern, indicating almost perfect alignment along the direction of the fiber. With cotton, on the other hand, the spots are much more diffuse showing that a somewhat random arrangement exists; it is probable that the crystallites form spiral layers making an angle of about 30° with the fiber length (R. O. Herzog, 1928). Many of the differences in physical properties, e.g., tensile strength, softness of fabric, penetration by water, etc., of cellulose fibers from various sources have been correlated with the change in the fiber structure.

Silk, Wool and Hair.—The natural fibers made of proteins, e.g., silk, wool and hair, do not give such sharp diffraction patterns as does cellulose; nevertheless, the photographs are definite enough to show a fiber structure and to permit the dimensions of the unit cell to be determined. For silk the data are as follows:

$$a = 9.68 \text{ Å.} \quad b = 7.00 \text{ Å.} \quad c = 8.80 \text{ Å.} \quad \beta = 75^\circ 50',$$

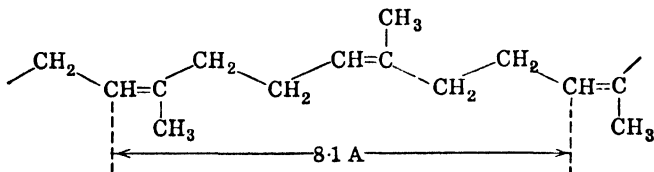
the b -axis again corresponding to the fiber length. The essential constituent of natural silk is the protein fibroin, which is known to consist of alanine and glycine residues; the fiber structure proposed by H. Mark and K. H. Meyer (1928), depicted below, shows an alternating series of such residues which give the correct cell dimensions. Each fiber consists of a number of chains of this type held approximately parallel to each other.



Observations of exceptional interest were made by W. T. Astbury (1930) on hair and wool. In the natural state these substances, and animal hairs generally, give almost the same X-ray diffraction pattern, attributed to oriented crystallites of the protein keratin. When the animal hairs are stretched, however, the positions of the spots are completely changed, although the X-ray photograph retains the characteristics of a fiber pattern, which now bears a striking resemblance to that for silk. In the unstretched or α -state the unit of structure repeats itself every 5.1 Å., but in the stretched or β -form the repetition occurs after 3.4 Å. Since the X-ray pattern of the latter form of keratin resembles that of fibroin, it is probable that β -keratin consists of a chain of amino-acid ($-\text{NHCOCHR}-$) residues somewhat similar to that shown above. In the natural hair or wool fiber the chains are linked by bridges of sulfur atoms, i.e., $-\text{S}-\text{S}-$, as in cystine, which can be obtained from keratin.

In fibroin each repeating unit consists of two amino-acid residues, and so the spacing is approximately twice that in β -keratin. The original views concerning the structure of α -keratin are now known to be unsatisfactory, and it is probable that this substance occurs in the form of a folded spiral, with three amino-acid residues to each turn of the spiral; this would account for the observed spacing of 5.1 Å. Upon exerting a tension on the α -keratin, the spiral presumably uncoils, thus leading to the long-chain β -form. It is of interest to mention that if the stretching load is removed β -keratin returns to the α -form, but if the fiber is exposed to steam while being stretched, a more or less permanent set is acquired in the extended form. However moisture and heat, without the application of tension, result in a tendency to return to the coiled structure. These observations have been used to account for such familiar phenomena as "permanent wave" of hair, and the creasing and shrinkage of woollen materials.

Rubber.—In the unstretched state rubber gives an X-ray diffraction pattern of diffuse rings indicative of a random structure, but when stretched a definite fiber structure is apparent (J. R. Katz, 1925); the length of the unit cell in the direction of the fiber is 8.1 Å., and the structure shown below has been proposed, taking into consideration the known fact that rubber is made up of a number of isoprene residues and has the formula $(\text{C}_5\text{H}_8)_n$.



When the tension is relieved rubber returns immediately to its amorphous state; it differs from wool in the respect that the latter has a regular fiber structure in both normal and stretched forms. It has been suggested (H. Mark, 1929) that normal rubber consists of a random arrangement of irregularly coiled

chains of isoprene units which straighten out on stretching but are able to return to their original form; the change in the X-ray pattern and the remarkable elasticity of rubber can be explained in this manner.

Slightly vulcanized rubber exhibits similar X-ray behavior to that of raw, i.e., natural, rubber, but when highly vulcanized no fiber structure appears upon stretching. It seems that vulcanization introduces bridges of sulfur atoms between the isoprene chains, somewhat similar to those existing in wool and hair. As the degree of vulcanization is increased, each chain is linked to so many others that the whole represents a three-dimensional, but random, network of coiled chains. Because of its relative rigidity, this network is less extensible than raw rubber, and there is considerable resistance to the alignment of the chains upon stretching.

In addition to the substances mentioned, investigations have been made of various natural and synthetic materials; the results for muscle fibers are of particular interest in relation to the problems of muscular action. Studies have also been made on artificial polymers, such as form the basis of synthetic plastics, and on the fibrous structures obtained when metals are drawn into wires. A discussion of these topics, is, however, outside the scope of this book.²²

Size of Crystallites.—It has been mentioned that the crystallites in fibers are generally much longer in the direction of the fiber length than across it; information of this nature is obtained by means of an equation deduced by M. von Laue (1926), which connects the breadth of the spots in the diffraction pattern with the dimensions of the crystalline particles. In cellulose the crystallites are approximately 500 Å. in length and about 50 Å. in thickness in both directions at right angles, so that each consists of a bundle of some twenty chains of about a hundred glucose residues. The chains in any one crystallite are probably all arranged in a definite manner, as in a space lattice, but the crystallites themselves have a random orientation with respect to each other, except for the fundamental fact that their lengths are all nearly parallel to the length of the fiber. In stretched rubber the crystallites are about 600 Å. long, 500 Å. wide and 150 Å. in thickness, so that they are virtually flat plates; they lie with their long axes parallel to the direction of stretching and their broader sides roughly parallel to the plane of the rubber film (J. Hengstenberg, 1928).

ENERGY OF IONIC CRYSTAL LATTICES

Attraction and Repulsion in a Crystal.—The stability of an ionic crystal depends on the balancing of at least three forces: (a) the electrostatic, or Coulomb, forces between the ions which give a resultant attraction falling off with the square of the distance; (b) van der Waals forces of attraction diminishing according to the seventh power of the distance; and (c) interatomic repulsive forces, falling off still more rapidly with distance (p. 300). The result of the attractive and repulsive forces is to lead to an equilibrium position of minimum potential energy, i.e., of greatest stability, as depicted in Fig. 87; the distance r_0 , corresponding to the minimum, represents the equilibrium value of the ionic separation in

the stable crystal lattice. At all temperatures there will, however, be some thermal oscillation about the equilibrium positions; in some cases rotation also occurs (p. 422).

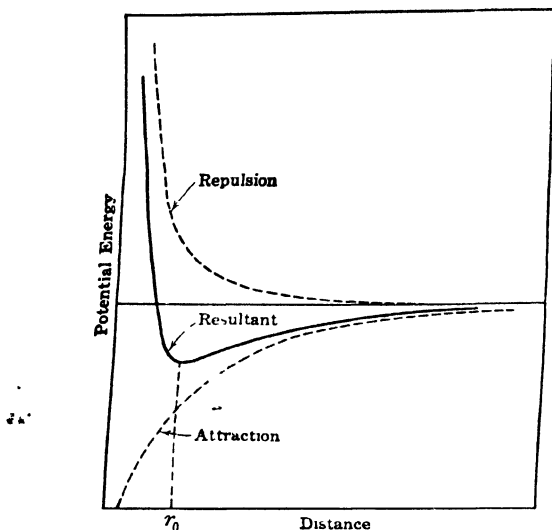


FIG. 87. Potential energy of crystal lattice

Assuming the potential energy of the ions to be zero when they are separated by an infinite distance, as indicated by the extension of the curve to the right, the minimum potential energy is equal numerically to the **lattice energy** of the crystal. This quantity is defined as *the decrease in energy accompanying the process of bringing the ions, when separated from each other by an infinite distance, to the positions they occupy in the stable lattice*. Another definition of the lattice energy is to regard it as the energy of formation of a given quantity of the crystal, e.g., 1 mole, from "gaseous ions," that is, from the separated ions free from any solvent. In this way it is seen that the lattice energy is the same in magnitude as, but opposite in sign to, the energy of dissociation of the crystal. It follows, therefore, that the greater the lattice energy, numerically, the greater the energy required to break up the crystal into its constituent ions.

Calculation of the Lattice Energy.—A method of calculating the lattice energy of an ionic crystal was proposed by M. Born and A. Landé (1918) which did not take into account the van der Waals forces. It has since been improved, but the original treatment, although approximate, will serve to illustrate the general principles. The potential energy per unit cell is represented by the following expression, in which the first term gives the Coulomb energy and the second the repulsion energy

$$\phi = -\frac{\alpha^2 e^2 n A}{r} + \frac{B}{r^2}, \quad (20)$$

where α is the largest common factor in the valences of the ions, e is the electronic charge, n the number of molecules per unit cell, and r is the distance between the ions. The quantity A , known as Madelung's constant, depends on the arrangement of the positive and negative ions in the crystal, and can be calculated by summing the mutual potential energies of all the ions in the lattice (E. Madelung, 1918); values have been worked out and tabulated for lattices of different types. The constant B , in the repulsive energy term, is determined by utilizing the fact that at the equilibrium distance r_0 the potential energy ϕ is a minimum, so that when $r = r_0$, $d\phi/dr$ is zero; it is readily seen that

$$B = \frac{\alpha^2 e^2 n A r_0^{-7}}{x}. \quad (21)$$

The other constant x involved in the repulsive energy term is evaluated from the compressibility of the crystal; for alkali halides it is found to vary from about 6 to 10, thus showing that the repulsive force falls off extremely rapidly with increasing distance. The potential energy per mole is obtained if ϕ is multiplied by the Avogadro number N and divided by n , the number of molecules in the unit cell, i.e., $N\phi/n$; the lattice energy per mole at equilibrium (U_0) is equal in magnitude but opposite in sign, so that

$$U_0 = -\frac{N\phi}{n} = \frac{\alpha^2 e^2 N A}{r_0} \left(1 - \frac{1}{n} \right). \quad (22)$$

The equilibrium distance r_0 is obtained either from X-ray measurements or, approximately, from the density of the crystal and so the lattice energy can be calculated provided the type of lattice is known.

In the later treatment (M. Born and J. E. Mayer, 1932) not only is the term c/r^6 included for the van der Waals energy (p. 299), but allowance is made for the zero-point energy (ϵ_0) of the molecules in the crystal; further, the repulsion term is replaced by the exponential expression $b e^{-r/\rho}$ as required by wave mechanics (p. 300). The equation for the lattice energy is then

$$U = \frac{\alpha^2 e^2 N A}{r} - b e^{-r/\rho} + \frac{c}{r^6} + \epsilon_0,$$

and utilizing the fact that dU/dr is zero when r is equal to r_0 , the equilibrium value is given with sufficient accuracy as

$$U_0 = \frac{\alpha^2 e^2 N A}{r_0} \left(1 - \frac{\rho}{r_0} \right) + \frac{c}{r_0^6} + \epsilon_0. \quad (23)$$

The constant c is obtained by the method of F. London, described in Chapter IV, and ρ is calculated from the compressibility. Values of the lattice energy have been determined by means of this equation, but the results do not differ very greatly from those given by the approximate method. In order to indicate the magnitude of the lattice energy the values for some alkali halides are quoted in Table 58.

TABLE 58. LATTICE ENERGY OF ALKALI HALIDES IN KCAL. PER MOLE

	Fluoride	Chloride	Bromide	Iodide
Sodium	216	184	176	164
Potassium	193	168	162	153
Rubidium	183	162	156	148

The Coulomb energy term is the most important factor in the lattice energy, and as this increases with the valence of the ion, so also does the lattice energy; further, the Madelung constant is approximately 2 for salts of the type AX, but is about 7 for the A_2X and AX_2 types, so that this also tends to increase the lattice energy for salts with ions of high valence.

Applications of Lattice Energy.—One of the most valuable applications of the lattice energy is to calculate the electron affinities of halogen atoms, i.e., the work required to remove an electron from a halogen ion, since these can only be determined directly with some difficulty. Use is made of the Born-Haber cycle (M. Born, 1919; F. Haber, 1919), which is based on the first law of thermodynamics. The crystalline halide MX is made to undergo theoretically the following sequence of changes:

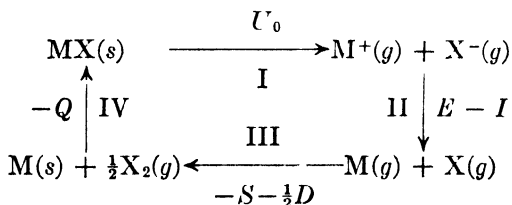
I. The crystal MX is broken up into gaseous ions M^+ and X^- . The energy absorbed* is the dissociation energy, and this is numerically equal to the equilibrium value of the lattice energy, i.e., U_0 .

II. An electron is removed from the X^- ion and added to the M^+ ion, leaving M and X as neutral *atomic* gases. If I is the ionization potential of the M atom (p. 50), and E is the electron affinity of the halogen atom X, then the energy involved is $E - I$.

III. The gaseous alkali metal is condensed to the solid state, and the halogen atoms X are combined to form X_2 molecules. The energy changes in these two processes are $-S$, where S is the heat of sublimation of the metal, and $-\frac{1}{2}D$, where D is the heat of dissociation of the molecular halogen, i.e., the heat absorbed when one molecule decomposes into two atoms.

IV. The alkali metal and the halogen molecules, in the forms they normally have at ordinary temperatures, i.e., in their standard states, are then combined to re-form the crystalline alkali halide MX; the heat absorbed is $-Q$, where Q is the thermochemical heat of formation of the salt.

The cycle which has thus been completed may be represented by the scheme



and hence it follows, by Hess's law, that

$$U_0 = Q + S + \frac{1}{2}D - E + I. \quad (24)$$

Utilizing the lattice energy (U_0), as calculated above, the known heats of

* Energy absorbed by the system is regarded as positive, and energy liberated is negative.

sublimation (S), dissociation (D) and formation (Q), and the ionization potential (I) of the metal, it is then possible, by means of (24), to evaluate the electron affinity (E) of the halogen. The result for any particular halogen, e.g., chlorine, may be obtained from any or all the alkali chlorides and the agreement of the different values is an indication of the reliability of the calculation; the chief source of error is the uncertainty concerning the lattice energies. Some of the results obtained in this manner are given in Table 59;* the electron affinities of the four halogen atoms are

TABLE 59. CALCULATION OF THE ELECTRON AFFINITY OF THE CHLORINE ATOM

Salt	U_0	Q	S	$\frac{1}{2}D$	I	E_{Cl}
LiCl	201	97	39	29	124	88 kcal.
NaCl	184	98	26	29	118	87
KCl	168	104	20	29	100	85
RbCl	162	105	19	29	96	87

found to be approximately as follows:

$$F = 95; \quad Cl = 87; \quad Br = 82; \quad I = 74 \text{ kcal.}$$

By comparing the rates of negative ion and electron emission from a heated surface in the halogen vapor, and in other ways, direct measurements have been made of the electron affinities; the mean experimental values are 86 ± 1 kcal. for chlorine, 84 ± 4 kcal. for bromine, and 73 ± 2 kcal. for iodine. In view of the difficulties involved in the measurements, the agreement with the results from lattice energies is very striking.

Since the electron affinities have been determined from the lattice energies, it should be possible to reverse the procedure and calculate the lattice energies of the halides of metals other than those of the alkali group, provided the quantities Q , S , D and I are available. The data obtained in this manner are of great importance. In the first place, they are reasonably accurate, for the calculation involves no assumption as to the nature of the units in the space lattice; and in the second place, by comparing these values with those obtained from (22) or (23), based on a rigid ionic lattice, there may be some indication of the type of linkage in the crystal. A few examples are given in Table 60 to show the agreement, and the significance of the discrepancies; the "experimental" values are those derived by the method of the Born-Haber cycle and the known electron affinities of the halogens, whereas the "theoretical" results have been evaluated by means of equation (23). In magnesium and calcium fluorides the linkages are undoubtedly ionic and the agreement between

* Ionization potentials are usually, and electron affinities sometimes, quoted in electron-volts (e.v.); the conversion factor to kcal. is derived as follows. The equivalent of 1 e.v. for a single molecule, atom or ion is 1.601×10^{-12} erg (p. 6), and so for 1 g. atom, g. ion or mole it is $1.601 \times 10^{-12} \times 6.023 \times 10^{23}$ erg. To convert into calories it is necessary to divide by 4.184×10^7 , giving 23,050 cal.; hence 1 e.v. = 23.05 kcal. per mole or per g. ion.

experimental and theoretical lattice energies is good. In the silver salts there is a tendency for polarization to occur, and the same tendency is very marked in cadmium iodide; the discrepancy between the experi-

TABLE 60. EXPERIMENTAL AND THEORETICAL LATTICE ENERGIES IN KCAL. PER MOLE

Salt	Experimental	Theoretical	Difference	Difference Per Cent
MgF ₂	689	697	+ 8	+ 1
CaF ₂	617	618	+ 1	0
AgCl	214	203	-11	- 5
AgBr	211	197	-14	- 7
AgI	208	190	-18	- 9
CdI ₂	563	474	-89	-16

mental lattice energy and that calculated on the assumption of purely ionic linkages increases as the valence partakes increasingly of covalent character.

Heats of Solution and Lattice Energies.—A solid salt MX may either be directly dissolved in excess of water to form hydrated ions, when the energy absorbed is equal to the total heat of solution, q_{MX} ; or it may first be split up into gaseous ions (the energy absorbed is U_{MX} , the lattice energy) and the ions then hydrated, absorbing energy $-(W_{M^+} + W_{X^-})$. It follows, therefore, that

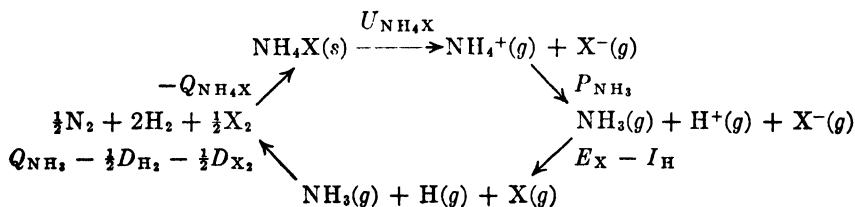
$$\begin{aligned} q_{MX} &= U_{MX} - (W_{M^+} + W_{X^-}); \\ \therefore U_{MX} - q_{MX} &= W_{M^+} + W_{X^-}. \end{aligned} \quad (25)$$

For another salt MY, having the cation the same as in MX, the corresponding equation is

$$\begin{aligned} U_{MY} - q_{MY} &= W_{M^+} + W_{Y^-}; \\ \therefore W_{X^-} - W_{Y^-} &= (U_{MX} - q_{MX}) - (U_{MY} - q_{MY}). \end{aligned} \quad (26)$$

The difference between $U - q$ for two salts having a cation is common should thus be independent of the nature of the anion; this rule has been verified by using experimental values of q , the integral (or total) heat of solution, and the lattice energies calculated by the method already described.

Proton Affinity.—The "proton affinities" of ammonia and of water are important quantities which may be estimated by a modification of the Born-Haber cycle; for ammonia, for example, the cycle which commences with a crystalline ammonium halide, viz., $NH_4X(s)$, may be written



where P_{NH_3} is the affinity of ammonia for a proton, i.e., the heat evolved when 1 mole of ammonia combines with protons; the other terms are self explanatory. From the data for the ammonium halides the mean value for P_{NH_3} of 207 kcal. has been derived.

The X-ray diffraction patterns of $\text{NH}_4 \cdot \text{ClO}_4$ and $\text{H}_3\text{O} \cdot \text{ClO}_4$, i.e., $\text{HClO}_4 \cdot \text{H}_2\text{O}$, are almost identical, and so it may be assumed that their lattice energies are equal. If the cycles, similar to that just given, are written out for these two substances, it is seen that they differ only in respect of the proton affinities P_{NH_3} and $P_{\text{H}_2\text{O}}$, the heats of formation of ammonia and water vapor, i.e., Q_{NH_3} and $Q_{\text{H}_2\text{O}}$, and the heats of formation of the two perchlorates, i.e., $Q_{\text{NH}_4 \cdot \text{ClO}_4}$ and $Q_{\text{H}_3\text{O} \cdot \text{ClO}_4}$; it can then be readily shown that

$$P_{\text{H}_2\text{O}} = P_{\text{NH}_3} + Q_{\text{NH}_3} - Q_{\text{H}_2\text{O}} + Q_{\text{H}_3\text{O} \cdot \text{ClO}_4} - Q_{\text{NH}_4 \cdot \text{ClO}_4}. \quad (27)$$

From the value of P_{NH_3} obtained above, J. Sherman (1932) calculated the proton affinity of water vapor to be 182 kcal. per mole.²³

HEAT CAPACITIES OF SOLIDS

In the deduction of lattice energies the special case was considered of a substance held together by ionic bonds, but the general conclusion that the constituent particles take up equilibrium positions of minimum potential energy, where the forces of attraction and repulsion exactly balance each other, is applicable to any crystalline solid irrespective of the nature of the bonds between the units. The atoms or ions are not at rest, but vibrate about their equilibrium positions, even at the absolute zero, and as the temperature is raised the amplitude of vibration increases, and hence so also do the mean potential and kinetic energies. It is the rate of increase of the vibrational energy with temperature that determines the heat capacity of the solid.

The Law of Dulong and Petit.—Before considering the subject from the theoretical standpoint, some of the experimental facts will first be examined. P. L. Dulong and A. T. Petit (1819) found that the product of the atomic weight and specific heat of solid elements was constant; that is to say, *all solid elements have the same heat capacity per g. atom, or the atoms of all elements have the same heat capacity.* Any of these statements may be taken as the expression of the **law of Dulong and Petit**; this law is not exact, but its approximate accuracy may be seen from the data in Table 61, where the specific heats are mean values for ordinary temperatures and atmospheric pressure.

TABLE 61. ATOMIC HEAT CAPACITIES OF SOLID ELEMENTS AT CONSTANT PRESSURE

Element	Atomic Weight	Specific Heat	Atomic Heat	Element	Atomic Weight	Specific Heat	Atomic Heat
Li	7	0.92	6.4 cal.	Ag	108	0.056	6.0 cal.
Al	27	0.21	5.7	I	127	0.052	6.6
Ca	40	0.15	6.0	Ta	181	0.033	5.9
Fe	56	0.11	6.2	Hg	200	0.033	6.6
As	75	0.083	6.2	U	238	0.027	6.4

The atomic heat capacities are almost constant at about 6.2 ± 0.4 cal. deg.⁻¹, in spite of the regular increase in the atomic weight from 7 to 238.

Although further work showed the law of Dulong and Petit to be at least approximately correct, H. V. Regnault (1841) and others brought to light the fact that the elements beryllium, boron, carbon and silicon have exceptionally low atomic heats, the values at ordinary temperatures being 3.5, 2.5, 1.35 (diamond) and 4.7 cal. deg.⁻¹, respectively. The atomic heat capacities of these elements increase rapidly with rising temperature, as the following results for diamond show:

Temperature	-50°	10.7°	85.5°	206°	615°	808° c.
Atomic Heat	0.76	1.135	2.11	3.28	5.33	5.44 cal. deg. ⁻¹

It was thought possible, therefore, that the law of Dulong and Petit would apply to all substances provided the temperature was high enough. This view is incorrect, however, for the heat capacities of even those elements which appear to obey the law at ordinary temperatures increase to some extent with increasing temperature, and for several elements values as high as 9 cal. are attained at their melting points. There is, consequently, no one set of conditions under which the Dulong and Petit law is applicable for all elements (W. A. Tilden, 1905).

Measurement of specific heats at low temperatures has shown that the atomic heats of all elements, and not only of those constituting the apparent exceptions to the law of Dulong and Petit, diminish with decreasing temperature. The temperature at which the rate of fall of atomic

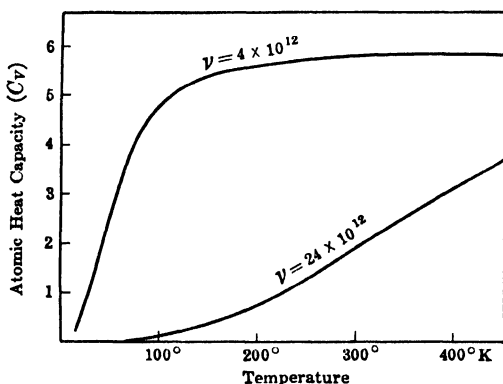


FIG. 88. Variation of atomic heat capacity with temperature. Upper curve, silver; lower curve, diamond

heat capacity becomes appreciable depends on the nature of the element, as may be seen in Fig. 88. It appears that at sufficiently low temperatures all solid elements will have zero specific and atomic heats; this would probably only be strictly true at the absolute zero, but at tempera-

tures below about -230°C . the atomic heat capacity of diamond is already too small to be measurable.

In view of the variation of atomic heat with temperature, it might be thought that the Dulong and Petit law is entirely fortuitous, although this is hardly likely since so many elements do actually have atomic heat capacities in the region of 6 cal. deg.^{-1} at ordinary temperatures. An examination of Fig. 88 shows that there is a definite tendency for the rate of increase of the atomic heat with temperature to slow down in the vicinity of the Dulong and Petit value, and it will be seen shortly that the atomic heat capacity of 6 cal. deg.^{-1} has a theoretical significance.

Molar Heat Capacities.—The work of Dulong and Petit was extended by F. E. Neumann (1831) and H. V. Regnault (1841) who observed that the products of the specific heat and molecular weight of solid compounds of similar composition were approximately constant. For example, the molar heat capacities of the series of carbonates RCO_3 , where R is magnesium, calcium, iron, zinc, barium or lead, are all very close to 21 cal. , and the values for the sulfates RSO_4 , where R is calcium, barium, strontium or lead, are nearly $21.5\text{ cal. deg.}^{-1}$ in each case. The significance of the results was first pointed out by J. P. Joule (1844) and subsequently by A. C. Woestyn (1848) and H. Kopp (1865): *the molar heat capacity of a solid compound is approximately equal to the sum of the atomic heat capacities of its constituents*. This conclusion is generally known as **Kopp's law**, although priority should be given to Joule; it may be put in the form

$$\text{Molar Heat Capacity} = n_1a_1 + n_2a_2 + n_3a_3 + \dots, \quad (28)$$

where n is the number of atoms of an element of atomic heat capacity a , the products being summed for all the elements in the molecule. Since for most elements at ordinary temperatures the atomic heat is about 6 cal. deg.^{-1} , it follows that

$$\text{Molar Heat Capacity} \approx 6n\text{ cal. deg.}^{-1}, \quad (29)$$

n being the total number of atoms in the molecule. Support for this rule is provided by the halides of many metals; for example, barium and strontium chlorides, lead chloride, bromide and iodide, and mercuric chloride and iodide all contain three atoms in the molecule and the molar heats are between 18 and 19 cal. deg.^{-1} . Similarly, the chlorides of silver, cuprous copper and potassium have molar heat capacities of 12 to 13 cal. deg.^{-1} , as required for molecules containing two atoms. The contribution of oxygen is usually only 4 cal. , and those of carbon and silicon are 1.8 and 3.8 cal. , respectively, in approximate agreement with their values in the elementary state. It is clear, therefore, that Kopp's rule of the additivity of the atomic heats of the elements in a compound is reasonably accurate.

Application of Atomic and Molar Heat Capacities.—The laws described above have been utilized for the purpose of fixing atomic weights

in a number of doubtful cases. The equivalent weight of uranium is 39.75, and as many of its compounds resemble those of iron, its valence was considered to be 3 and its atomic weight was taken as 119.25; Mendeléeff, however, found no room for such an element in his periodic classification and so proposed to adopt the double value, viz., 238.5. This figure was confirmed when the specific heat of metallic uranium was determined to be 0.027, so that the atomic weight must be approximately 240. The atomic weight of indium was verified in an analogous manner; the specific heat of the element is 0.057, and so the atomic weight should be about 110. The accurate value is thus three times the equivalent weight 37.8, i.e., 112.4.

Kopp's rule was employed by S. Cannizzaro (1858) to show that the atomic weight of mercury was 200, and that the value previously assumed, namely 100, was inaccurate. The specific heat of mercuric chloride HgCl_2 is 0.069, and if the atomic weight of mercury is 200, the molecular weight of the chloride is 271 and its molar heat capacity is 18.7 cal. deg.⁻¹, in accordance with expectation for a compound containing three atoms.

Classical Derivation of Heat Capacity.—An ideal solid may be assumed to consist of a space lattice of *independent* atomic units vibrating about their respective equilibrium positions, but not interacting with one another in any way. A gram atom of the crystal may thus be regarded as consisting of $3N$ harmonic oscillators, where N , the Avogadro number, is equal to the number of individual atoms, each of which can vibrate in three directions. According to the classical principle of the equipartition of energy (p. 329), each mode of vibration should contribute kT to the energy of an atom or molecule, where k is the Boltzmann constant, i.e., R/N . The total energy of the $3N$ oscillators is thus equal to $3NkT$, so that for 1 g. atom

$$E = 3RT. \quad (30)$$

Differentiation of E with respect to temperature at constant volume, that is $(\partial E/\partial T)_V$, gives the heat capacity C_V per g. atom of solid (cf. p. 189); hence,

$$C_V = 3R = 5.96 \text{ cal. deg.}^{-1}. \quad (31)$$

It follows, therefore, that the heat capacity of an ideal solid should be 5.96 cal. deg.⁻¹ for each g. atom present (L. Boltzmann, 1871). The nearest approach to ideal behavior should be found in monatomic solids; of these the metallic elements are the most familiar examples, and such substances should have an atomic heat, at constant volume, of about 6 cal. deg.⁻¹. It is evident that this conclusion provides a theoretical basis for the law of Dulong and Petit, and of the experimental fact that most elements have an atomic heat capacity of 6.4 cal. deg.⁻¹. The measurements are usually made at constant pressure, but the correction to constant volume may be made by the thermodynamic equation (p. 235)

$$C_P - C_V = TV\alpha_t^2/\beta, \quad (32)$$

where α , and β are the coefficients of thermal expansion and compressibility, respectively, and V is the atomic volume. Alternatively, the empirical equation of F. A. Lindemann and A. Magnus (1910)

$$C_P - C_V = aT^{\frac{1}{2}}, \quad (32)$$

where a is a constant for each substance, may be employed. The conversion to constant volume has the effect, as shown by G. N. Lewis (1907), of reducing the atomic heat capacities of many metals to values of 5.9 ± 0.2 cal. deg.⁻¹, although certain elements, particularly the alkali metals, have higher atomic heats. Some of the results are given in Table 62.

TABLE 62. ATOMIC HEAT CAPACITIES AT CONSTANT VOLUME

Al 5.7	Zn 5.6	Sr 6.1	Au 5.9
Fe 5.9	Pd 5.9	Sb 5.9	Tl 6.1
Ni 5.9	Ag 5.8	I 6.0	Pb 5.9
Cu 5.6	Cd 5.9	Pt 5.9	Bi 6.2

In a simple compound it is possible that each element will exert little influence on the oscillation of the others; each atom will then have the same heat capacity in the compound as in the elementary state. The molar heat capacity of a compound should then be approximately equal to $6n$ cal. deg.⁻¹, where n is the number of atoms in the molecule; this is the basis of Kopp's law. In actual practice the existence of chemical bonds constitutes a form of interference with the independent vibrations of the atoms, and instead of the oscillations being random there are certain preferred directions; in addition, other forms of internal energy, e.g., rotational and electronic, must be considered, and so Kopp's law can only be of limited applicability.

Quantum Theory of Specific Heats: The Einstein Equation.—The classical theory given above not only requires the atomic heat capacity to be $3R$ per g. atom, but it should be independent of temperature; all specific heats increase with temperature, however, and the effect is most marked for the light elements, whose heat capacities are exceptionally low (p. 414). Even the atomic heats of those elements having values of about 6 cal. deg.⁻¹ at normal temperatures diminish at low temperatures, so that it is partly fortuitous that they behave in accordance with classical requirements. An important step was taken by A. Einstein (1907) when he applied the quantum theory to the problem of the heat capacity of solids. According to Planck (p. 33), the mean energy of vibration $\bar{\epsilon}$ of a linear oscillator, i.e., oscillating in one direction in space, in excess of the zero-point value, is given by

$$\bar{\epsilon} = \frac{h\nu}{e^{h\nu/kT} - 1}, \quad (34)$$

where k is the Boltzmann constant and ν is the characteristic frequency of the oscillator. As seen above, 1 g. atom of an element can, however,

be regarded as a system of $3N$ linear oscillators, N being the Avogadro number. The energy per g. atom is then

$$E = 3N \frac{h\nu}{e^{h\nu/kT} - 1}, \quad (35)$$

and Einstein's equation for the heat capacity at constant volume is

$$C_V = \left(\frac{\partial E}{\partial T} \right)_V = 3R \left(\frac{h\nu}{kT} \right)^2 \frac{e^{h\nu/kT}}{(e^{h\nu/kT} - 1)^2}. \quad (36)$$

According to this relationship C_V should approach zero at very low temperatures, whereas at high temperatures, when $h\nu/kT$ is small in comparison with unity, the value becomes $3R$, in agreement with the classical treatment and the law of Dulong and Petit. Under the latter conditions an appreciable number of vibrational levels are occupied (cf. p. 337), and the atom behaves as if it were a classical oscillator. For most elements $h\nu/kT$ is sufficiently small at ordinary temperatures for the atomic heat capacity to be about 6 cal. deg.⁻¹, but if ν is large, that is, when the atoms are very firmly bound, e.g., as in diamond, a much higher temperature is necessary before the classical heat capacity is attained. In general, it can be concluded that those elements which constitute exceptions to the law of Dulong and Petit have high values of ν , the characteristic frequency of the atomic oscillators.

The only variable in the Einstein equation is ν , and the plot of C_V against T , according to (36), is of the type found by experiment (cf. Fig. 88); the exact shape depends on the value of ν which must be obtained empirically.

The Debye Equation.—The Einstein equation is obviously a step in the right direction, but closer examination has shown that at low temperatures the specific heats determined by it fall off more rapidly than do the experimental values. An attempt was made by W. Nernst and F. A. Lindemann (1911) to improve (36) empirically; a formula was suggested involving two terms of the type of the Einstein equation, one with $\nu' = \nu/2$, and the other with $\nu'' = 2\nu$. The new empirical relationship gave very good agreement with experiment, but it was rapidly superseded by the theoretical improvements made by P. Debye (1912), and independently, but less simply, by M. Born and T. von Kármán (1912). The Einstein treatment, like the classical one of Boltzmann, implies a lattice of independent atoms; even in a monatomic solid, however, this cannot be strictly true. It is not correct, therefore, that all the N atoms in 1 g. atom of a crystal oscillate at the same frequency, and a more reasonable postulate would be that the g. atom involves a *coupled system* of $3N$ oscillators; that is,

$$E = \sum_{r=1}^{3N} \frac{h\nu_r}{e^{h\nu_r/kT} - 1}, \quad (37)$$

where ν_r is the frequency for a particular value of r which can vary from 1 to $3N$. In order to evaluate this expression Debye assumed the distribution of frequencies to be the same as that of sound waves in a homogeneous, isotropic elastic medium. By means of the theory of elasticity the number of such vibrations per unit volume (dz) lying between the frequencies ν and $\nu + d\nu$ is found to be

$$dz = 9N \frac{\nu^2 d\nu}{\nu_m^3}, \quad (38)$$

where ν_m represents the maximum of the $3N$ vibration frequencies. If this result is introduced into (37), and the summation replaced by integration, which is justifiable in view of the large number of frequencies and their probable regular distribution, the relationship

$$E = \frac{9N}{\nu_m^3} \int_0^{\nu_m} \frac{h\nu}{e^{h\nu/kT} - 1} \nu^2 d\nu \quad (39)$$

is obtained. On differentiation with respect to temperature, the Debye heat capacity equation

$$C_V = \frac{9R}{\nu_m} \int_0^{\nu_m} \frac{e^{h\nu/kT}}{(e^{h\nu/kT} - 1)^2} \left(\frac{h\nu}{kT} \right)^2 \nu^2 d\nu \quad (40)$$

results. The quantity $h\nu/k$ has the dimensions of temperature, and the special value $h\nu_m/k$ for each substance is called its **characteristic temperature**; it is represented by the symbol θ . Further, $h\nu/kT$ is dimensionless and may be replaced by the variable x ; (40) then becomes

$$C_V = 3R \left[\frac{12T^3}{\theta^3} \int_0^{\theta/T} \frac{x^3 dx}{e^x - 1} - \frac{3\theta/T}{e^{\theta/T} - 1} \right]. \quad (41)$$

A method has been devised for evaluating the integral, and the results have been tabulated for various values of θ/T . In order to determine the heat capacity of an element at any temperature it is only necessary, therefore, to know the value at any one temperature so that θ can be determined, and then by the use of tables based on (41) the heat capacity at any other temperature can be obtained directly. It is possible in this manner to test the reliability of the Debye equation and excellent agreement has been found with experimental data. The required results can also be obtained by a graphical method, as will be seen shortly.

At very low temperatures θ/T is large, and by putting it equal to infinity the integral in (41) can be evaluated; at the same time the last term becomes negligible. The Debye equation for low temperatures is then

$$C_V = \frac{1}{5} \pi^4 R \left(\frac{T}{\theta} \right)^3 = 464.4 \left(\frac{T}{\theta} \right)^3, \quad (42)$$

or

$$C_V = aT^3, \quad (43)$$

where a is a constant for the particular element. The fact that the heat capacity, at constant volume, should be proportional to the cube of the absolute temperature, at low temperatures, is one of the most important consequences of the Debye theory of the specific heats of solids. It has been verified not only for monatomic solids, but also for a few simple compounds crystallizing in the cubic system.*

If the temperature is relatively high, both x and θ/T are small; the Debye equation then reduces to the classical form, $C_V = 3R$, so that the heat capacities of elements should be about 6 cal. deg.⁻¹ g. atom⁻¹. The higher values observed, for the alkali metals in particular, cannot be explained either by the Debye or the Einstein theory. It has been suggested that this is due to contributions to the energy made by the electrons.

According to the Debye equation (41), and also to the Einstein equation (36), if ν is assumed to be equal to ν_m of the former, C_V is a definite function of T/θ for all monatomic elements, i.e.,

$$C_V = f(T/\theta), \quad (44)$$

where the value of θ depends on the element. In other words, if C_V is plotted against θ/T or T/θ , or $\log (T/\theta)$, the same curve should apply to all solid elements; that this is true is demonstrated by Fig. 89, where it is

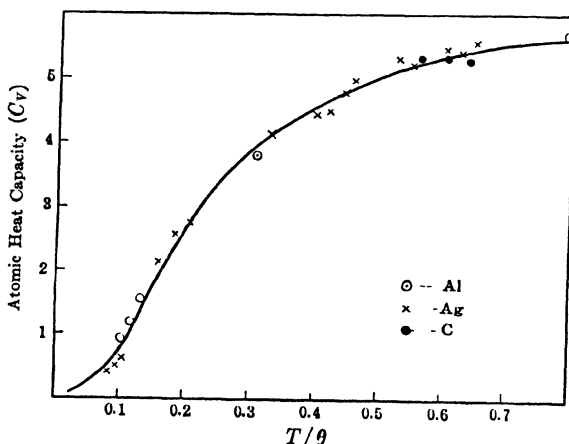


FIG. 89. Atomic heat capacity as a function of the characteristic temperature

seen that the experimental data for substances as diverse as silver and diamond fall on the one curve. Once the form of this curve has been established it is only necessary to know C_V at one temperature, preferably on the rising part of the curve where the value is well below $3R$, for the exact position of the curve for any particular element to be definitely fixed; in this manner the variation of heat capacity over a considerable range of temperature can be determined.

* For methods of measuring heat capacities at low temperatures, see p. 866.

It should be noted that the Debye specific heat equation was only meant to apply to isotropic solids, that is, those belonging to the cubic system (see p. 346); many metallic elements do in fact satisfy this requirement and so they obey the equation. It appears that the equation is also applicable to some metals, e.g., zinc, crystallizing in the hexagonal system. For other elements, and some compounds, G. N. Lewis and G. E. Gibson (1917) put forward the more general form

$$C_V = f(T/\theta)^n, \quad (45)$$

where f represents the Debye function and n is less than unity for substances to which the Debye equation does not apply. By giving an appropriate value to n the relationship has been found to hold for graphite, rhombic sulfur, iodine and for a number of compounds, mainly halides of metals.

Mention may be made of other empirical heat capacity equations which have been applied to binary compounds. W. Nernst (1914) proposed to express C_V as the weighted sum of an Einstein (f_E) and a Debye (f_D) function; thus,

$$C_V = af_E(T/\theta_1) + bf_D(T/\theta_2), \quad (46)$$

where θ_1 and θ_2 are constants for the given compound. Alternatively, the sum of two Debye functions may be employed. These empirical equations, as well as that of Lewis and Gibson, are often of value when it is required to express the variation of heat capacity of a compound with temperature, especially in connection with the determination of entropy (see Chapter XI).²⁴

Calculation of Characteristic Temperature.—The only quantity in the Debye treatment that is dependent on the nature of the metal is the characteristic temperature θ , which is related to the frequency ν_m (p. 419). The value of θ can, of course, be derived from a measurement of the heat capacity, as explained above, but it is of interest to see if it can be calculated from other properties of the solid. Debye showed that ν_m (or θ) should be related to the elastic constants of the medium, viz., the compressibility and Poisson's ratio; in the few cases where the necessary data are available, the calculated values of the characteristic temperature have been found to be in good agreement with the experimental (heat capacity) results (see Table 63).

As an approximation it may be supposed that the Debye and Einstein frequencies are the same; the latter is given by the expression (A. Einstein, 1911)

$$\nu \approx 2.5 \times 10^7 v^{1/6} / M^{1/3} \beta^{1/2}, \quad (47)$$

where ν is the specific volume, M the atomic (or molecular) weight and β is the compressibility. A semi-empirical equation, based partly on experimental values of the Einstein frequency, and partly on the supposition (F. A. Lindemann, 1910) that at the melting point (T_m) the amplitudes of vibration of the atoms in a solid are approximately equal to their mean distance apart, is

$$\nu \approx 2.8 \times 10^{12} \sqrt{T_m / M V^{2/3}}, \quad (48)$$

where M is the atomic (or molecular) weight and V is the atomic (or molecular) volume. Values of the characteristic temperature ($\theta = h\nu/k$), calculated from the frequencies given by (47) and (48), as well as those derived from elasticity

data, are compared with those obtained experimentally from heat capacity measurements, in Table 63.

TABLE 63. CHARACTERISTIC TEMPERATURES OF SOLID ELEMENTS

Element	From Heat Capacity	From Elasticity	From Compressibility	From Melting Point
Lead	88	73	103	85
Cadmium	150	—	124	133
Silver	215	214	195	210
Copper	315	332	270	330
Aluminum	398	402	314	360
Iron	453	—	310	395
Carbon	1860	—	—	1570

It will be seen from (47) and (48) that elements of low compressibility, i.e., hard elements, and of high melting point should give the highest values of θ , and hence elements with such properties should only attain the classical atomic heat of 6 cal. deg.⁻¹ at relatively high temperatures. The elements carbon (diamond), boron and silicon provide excellent confirmation of this deduction. The low atomic weight of these substances is, of course, a contributory factor, as may be seen from the equations mentioned, but it is not the most important; lithium, for example, has a low atomic weight, but its atomic heat capacity attains a value of 6 cal. deg.⁻¹ at room temperature because it is a soft metal of low melting point.

ROTATION IN THE SOLID STATE

The Possibility of Rotation.—It was suggested by L. Pauling (1930) that certain polymorphic changes in crystals, accompanied by the absorption of heat, could be explained on the assumption that the molecules or ions constituting the lattice commenced to rotate freely at a particular temperature when the crystal was heated. At low temperatures the energy of rotation may be insufficient to overcome the lattice forces, and the molecule will undergo an oscillatory, or to and fro, rotation, i.e., **libration**. When the thermal energy, however, exceeds the potential energy of the lattice *free* rotation can set in. The temperature at which this occurs will generally be sharp, like a melting point, because the rotation of one molecule results in a weakening of the molecular forces in its neighborhood, and so rotation of the other molecules is facilitated. The ease with which rotation occurs depends on the magnitude of the attractive forces within the crystal and also on the lattice dimensions; in some instances the latter are such as to make rotation of a relatively large molecule almost impossible. For many substances, therefore, rotation cannot set in at any temperature below the melting point, and so does not occur in the solid state.

Evidence for Rotation.—The most direct methods for detecting the onset of rotation are based on the study of crystal structure and of dielectric constant, although the variation of specific heat and of thermal expansion with

temperature also provide valuable information. It is to be expected that rotation will be accompanied by an increase in the symmetry of the crystal lattice, and it is found that rotation occurs mainly in the cubic, and to some extent in the hexagonal, system; both of these can be built up from close packed spheres (p. 378). For example, at -174.8°C . crystalline hydrogen chloride undergoes a polymorphic transformation from orthorhombic to cubic forms; in the latter, which is stable above the transition point, there is free rotation about an axis perpendicular to the line through the centers of the atoms. Definite proof that the transition corresponds to the change from libration to free rotation is provided by measurement of the dielectric constant. Hydrogen chloride is a polar molecule, and provided rotation can occur about an axis at right angles to the direction of the dipole, the dielectric constant should be high, whereas inability to rotate in this manner will lead to a low dielectric constant. The change in form of hydrogen chloride to cubic crystals is accompanied by a sharp increase in the dielectric constant of the solid, showing that rotation occurs at temperatures above -174.8°C . The transition from tetragonal to cubic forms which ammonium nitrate undergoes at 125°C . is accompanied by the onset of free rotation of the NO_3^- ion. The change from libration to free rotation is not always accompanied by a change of crystalline form. In sodium nitrate, for instance, complete rotation of the NO_3^- ion occurs above 275°C ., but there is no alteration in crystal structure at this temperature. The transition is, however, accompanied by an absorption of heat and a change in thermal expansion. It is of interest that neither of these occurs suddenly, and the results indicate that there is an increase of rotational energy below 275°C . until at this temperature it is just sufficient to overcome the potential energy of the lattice and permit of free rotation of the nitrate ion. A variation in the rate of heat absorption obviously means a change in specific heat, and abnormalities of this kind have been observed to correspond to the onset of unrestricted rotation. It may be mentioned that solid ammonium chloride, bromide and iodide, belonging to the cubic system, exhibit transitions which do not correspond to any alteration in crystalline form; these occur at -30.4° , -38° and -42.5°C ., respectively, and above these temperatures free rotation of the ammonium ion takes place in each case. No rotation has been observed in ammonium fluoride; the lattice energy is probably exceptionally high because of the presence of hydrogen bonds (cf. p. 387).

With small molecules, having low or zero dipole moments, e.g., H_2 , D_2 , O_2 , N_2 , CO and CH_4 , rotation sets in at low temperatures; in these cases the dielectric constant cannot provide any information and the evidence for rotation must be obtained from specific heat and other thermal measurements. From the entropies of hydrogen and deuterium there is reason to believe that rotation still occurs in the solids at the lowest temperatures at which the specific heats have been measured (p. 876). With polar molecules rotation does not always set in so readily; for example, although methane rotates above -252.6°C ., ammonia, water and hydrogen fluoride show very little, if any, rotation in the solid state. Partly owing to the strong forces between the dipolar molecules in the crystal and partly, perhaps, because of hydrogen bonding, the lattice energy is high and the thermal energy necessary to permit rotation to take place is not attained while the substances are yet solid. From dielectric constant and thermal measurements, however, it is clear that the homologues of these compounds, viz., phosphine, hydrogen sulfide and hydrogen chloride, are all

capable of undergoing rotation above transition temperatures of -243° (?), -169.6° and -174.8° C., respectively. With the next homologues, arsine, hydrogen selenide and hydrogen bromide, rotation sets in at even lower temperatures. There is no evidence of rotation in solid chlorine, bromine and iodine, neither is there in nitric oxide, carbon dioxide, nitrous oxide and hydrogen cyanide. The absence of free rotation in nitric oxide is attributed to the formation of N_2O_2 molecules at low temperatures, and in the other cases the arrangement of molecules in the lattice presumably does not permit rotation about an axis perpendicular to the length of the molecule.

A number of organic compounds which do not depart appreciably from spherical symmetry, at least dimensionally, are able to rotate in the solid. As seen above, methane rotates in this manner, and so also do the molecules of tetramethyl methane, carbon tetrachloride, and the three methyl chloromethanes. Because the chlorine atom has approximately the same radius as the methyl group, the latter are virtually spherically symmetrical. Similar considerations apply to the tertiary butyl halides which also rotate in the solid state. Methyl and chloro-, penta and hexa, substituted benzenes are capable of rotation in the plane of the ring, as might be expected. In a number of long-chain compounds, e.g., hydrocarbons and primary alcohols, which are known to exist in α - and β -forms, rotation can take place about the axis of the chain in the high temperature, i.e., α -, form. In the β -forms of the alcohols there are two spacings of 3.7 and 4.2 Å. corresponding to the cross sections of the chains (cf. p. 399), but when rotation sets in there is only one spacing, viz., 4.2 Å., as is to be expected from the tendency toward increasing symmetry.²⁵

REFERENCES

- Astbury, W. T. "Fundamentals of Fibre Structure."
 Bragg, W. H. "Introduction to Crystal Analysis."
 Bragg, W. H. and W. L. "X-Rays and Crystal Structure."
 Bragg, W. L. "The Crystalline State," Vol. I.
 Bragg, W. L. "The Atomic Structure of Minerals."
 Buerger, M. J. "X-Ray Crystallography."
 Clark, G. L. "Applied X-Rays."
 Compton, A. H. and Allison, S. K. "X-Rays in Theory and Experiment."
 Davey, W. P. "A Study of Crystal Structure and its Applications."
 Evans, R. C. "Crystal Chemistry."
 James, R. W. "X-Ray Crystallography."
 Pauling, L. "The Nature of the Chemical Bond."
 Randall, J. T. "The Diffraction of X-Rays and Electrons, etc."
 Stillwell, C. W. "Crystal Chemistry."
 Wyckoff, R. W. G. "The Structure of Crystals."
 Zachariasen, W. H. "Theory of X-Ray Diffraction in Crystals."

1. Friedrich and Knipping, *Ann. Phys.*, **41**, 971 (1912).
2. W. L. Bragg, *Proc. Camb. Phil. Soc.*, **17**, 43 (1913); *Proc. Roy. Soc.*, **89A**, 248, 468 (1913).
3. W. L. Bragg, *Phil. Trans. Roy. Soc.*, **215**, 253 (1915); *Proc. Roy. Soc.*, **123A**, 537 (1929); Robertson, *ibid.*, **140A**, 79 (1933); *J. Chem. Soc.*, 1195 (1936); 219 (1937).
4. Patterson, *Phys. Rev.*, **46**, 372 (1934); *Z. Krist.*, **70**, 517 (1935); Harker, *J. Chem. Physics*, **4**, 381 (1936); Cox, *et al.*, *J. Chem. Soc.*, 883 (1937).
5. W. H. Bragg, *Proc. Roy. Soc.*, **88A**, 428 (1913).

6. W. L. Bragg, ref. (2); W. H. Bragg and W. L. Bragg, *Proc. Roy. Soc.*, **89A**, 246, 430 (1913).
7. Wyckoff, *J. Am. Chem. Soc.*, **42**, 1100 (1920); *op. cit.*
8. Polanyi and Weissenberg, *Z. Phys.*, **9**, 123; **10**, 144 (1922); Weissenberg, *ibid.*, **23**, 229 (1924); Schiebold, *ibid.*, **28**, 355 (1924); Bernal, *Proc. Roy. Soc.*, **113A**, 117 (1926); de Lange, Robertson and Woodward, *ibid.*, **171A**, 398 (1930).
9. Debye and Scherrer, *Phys. Z.*, **17**, 277 (1916); Hull, *Phys. Rev.*, **10**, 661 (1917).
- For general review of X-ray methods, see Warren, *J. App. Phys.*, **12**, 375 (1941).
10. Hume-Rothery, *J. Inst. Metals*, **35**, 313 (1926); "The Metallic State"; "The Structure of Metals and Alloys"; Westgren, *et al.*, *Phil. Mag.*, **50**, 331 (1925); *Trans. Farad. Soc.*, **25**, 380 (1929); Bradley, *Proc. Roy. Soc.*, **112A**, 678 (1926); **115A**, 456 (1927); Jones, *ibid.*, **144A**, 255 (1934); **147A**, 396; Weibke, *Z. Elek.*, **44**, 209, 263 (1938); Laves and Walbaum, *Z. anorg. Chem.*, **250**, 110 (1942); Pauling, "The Nature of the Chemical Bond," Chap. XI.
11. Bragg and Williams, *Proc. Roy. Soc.*, **151A**, 540 (1935); Nix and Shockley, *Rev. Mod. Phys.*, **10**, 1 (1938); see also *Ann. Rep. Chem. Soc.*, **32**, 185 (1935).
12. Pauling and Huggins, *Z. Krist.*, **87**, 205 (1934); Schomaker and Stevenson, *J. Am. Chem. Soc.*, **63**, 37 (1941); Pauling, ref. (10), Chaps V and X.
13. W. L. Bragg, *Phil. Mag.*, **2**, 258 (1926); Pauling, *Proc. Roy. Soc.*, **114A**, 181 (1927); *J. Am. Chem. Soc.*, **49**, 765 (1927); Goldschmidt, *Trans. Farad. Soc.*, **25**, 253 (1929).
14. Goldschmidt, ref. (13); see also, Wells, *Phil. Mag.*, **32**, 106 (1941).
15. Bragg, *Trans. Farad. Soc.*, **25**, 291 (1929); Stillwell, *op. cit.*
16. Pauling, *J. Am. Chem. Soc.*, **51**, 1010 (1929); Keggins, *Proc. Roy. Soc.*, **144A**, 75 (1934).
17. Dickinson, *J. Am. Chem. Soc.*, **44**, 276, 774, 2401 (1922); Cox, *et al.*, *J. Chem. Soc.*, 1912 (1932); **129**, 775 (1936); Powell *et al.*, *ibid.*, 359 (1935); **119** (1937).
18. Beevers and Lipson, *Proc. Roy. Soc.*, **146A**, 570 (1934); **148A**, 664 (1935); *Z. Krist.*, **91**, 157 (1935); Lipson, *Phil. Mag.*, **19**, 887 (1935); *Proc. Roy. Soc.*, **151A**, 347 (1935).
19. Cox, *J. Chem. Soc.*, 1635 (1936); Mulligan and Weiser, *J. Phys. Chem.*, **41**, 1029 (1937).
20. De Broglie and Friedel, *Compt. rend.*, **176**, 738 (1923); Piper and Grindley, *Proc. Phys. Soc.*, **35**, 269 (1923); Muller and Shearer, *J. Chem. Soc.*, **123**, 2043, 3156 (1923); Shearer, *ibid.*, p. 3152; Muller and Saville, *ibid.*, **127**, 599 (1925); Saville and Shearer, *ibid.*, p. 591; Gibbs, *ibid.*, **125**, 2622 (1921); Muller, *Proc. Roy. Soc.*, **120A**, 437 (1928); Malkin *et al.*, *ibid.*, **128A** (1930); **158A**, 691 (1937); *J. Chem. Soc.*, 3739 (1930); 2796 (1931), 666 (1931); 1628 (1936); 1409 (1937); Hendricks, *Chem. Rev.*, **7**, 431 (1930).
21. Lonsdale, *Proc. Roy. Soc.*, **123A**, 494 (1929); **133A**, 536 (1931); Robertson, *Chem. Rev.*, **16**, 417 (1935).
22. For reviews, see Astbury, *Ann. Rep. Chem. Soc.*, **28**, 322 (1931); *Phil. Trans. Roy. Soc.*, **230A**, 75 (1931); *op. cit.*; Stillwell, *op. cit.*; *J. Chem. Ed.*, **11**, 159 (1934); Randall, *op. cit.*; Clark, *op. cit.*; Taylor, *Proc. Am. Phil. Soc.*, **85**, 1 (1941); Fuller, *Chem. Rev.*, **26**, 143 (1940); Mark, *ibid.*, p. 169; Gehman, *ibid.*, p. 203; Corey, *ibid.*, p. 227; Huggins, *ibid.*, **32**, 195 (1943).
23. Lewis, *Ann. Rep. Chem. Soc.*, **17**, 1 (1920); Sherman, *Chem. Rev.*, **11**, 93 (1932); Born and Mayer, *Z. Phys.*, **75**, 1 (1932); Mayer and Helmholz, *ibid.*, p. 19; *J. Chem. Phys.*, **2**, 245 (1934); Mayer, *et al.*, *ibid.*, **1**, 270 (1933); **3**, 20 (1935); **8**, 282 (1940); **11**, 516 (1943); Huggins, *ibid.*, **5**, 113 (1937); Glockler and Calvin, *ibid.*, **3**, 371 (1935); **4**, 492 (1936); Fajans, *Z. phys. Chem. (Bodenstein Festband)*, 807 (1931); Verwey and de Boer, *Rec. trav. chim.*, **55**, 443 (1936); Baughan, *J. Chem. Soc.*, 1403 (1940); Briggleb, *Naturwiss.*, **30**, 469 (1942); see also, Rice, "Electronic Structure and Chemical Binding," Chap. XIV; Pauling, ref. (10), Chap. X.

24. Einstein, *Ann. Phys.*, **22**, 180 (1907); Debye, *ibid.*, **39**, 789 (1912); Born and Kármán, *Phys. Z.*, **13**, 294 (1912); **14**, 15 (1913); Lewis and Gibson, *J. Am. Chem. Soc.*, **39**, 2554 (1917); see also, Fowler and Guggenheim, "Statistical Thermodynamics," Chap. IV; Glasstone, "Theoretical Chemistry," Chap. VII.
25. Pauling, *Phys. Rev.*, **36**, 430 (1930); Fowler, *Proc. Roy. Soc.*, **149A**, 1 (1935); Smyth, *et al.*, *Chem. Rev.*, **19**, 329 (1936); *J. Am. Chem. Soc.*, **61**, 1695 (1939); **62**, 949, 2468 (1940); White, Biggs and Morgan, *ibid.*, p. 16; Eucken, *Z. Elek.*, **45**, 126 (1939); Biggs, *J. Chem. Ed.*, **18**, 224 (1941).

CHAPTER VI

CHANGES OF STATE¹

LIQUEFACTION OF GASES

Introduction.—J. B. van Helmont (died 1644), the inventor of the word “gas,” realized that, as the result of cooling, some gases could be condensed to liquids; these he referred to as “vapors,” to distinguish them from true gases. Systematic work on the liquefaction of gases did not commence, however, until the beginning of the nineteenth century, when the subject was taken up by Michael Faraday. It is reported that just previously sulfur dioxide and ammonia had been liquefied, and in 1805 liquid chlorine and hydrogen chloride were obtained by compressing the gases (T. Northmore). In 1823 Faraday described the liquefaction of hydrogen sulfide, cyanogen, carbon dioxide, nitrous oxide, and hydrogen bromide, in addition to the gases already mentioned; his apparatus consisted of a closed tube, similar to an inverted V in shape, into one limb of which was sealed a substance, or substances, capable of giving off a particular gas on warming or mixing. The other sealed limb was placed in a freezing mixture of ice and salt, so that when the gas was evolved it was subjected to both high pressure and low temperature; the resulting liquid was condensed in the cooled portion of the tube. At the time of Faraday’s earlier investigations the lowest known temperatures were obtained from mixtures of ice and salt; the introduction by M. Thilorier (1835) of a freezing mixture of solid carbon dioxide and ether, whereby temperatures as low as -110°C . could be obtained, therefore marked a considerable advance. Thilorier obtained liquid carbon dioxide in relatively large amounts by means of a wrought-iron apparatus operating on the same principle as that used by Faraday for small scale work; by partial evaporation of the liquid the temperature was lowered sufficiently for the remainder to solidify, and so solid carbon dioxide was obtained for the first time. With the aid of Thilorier’s mixture as a cooling agent, Faraday (1845) succeeded in liquefying ethylene, phosphine, silicon tetrafluoride and boron trifluoride, and even in solidifying a number of the more condensable gases. In spite of the low temperature attained, hydrogen, oxygen, nitrogen, nitric oxide, carbon monoxide and methane could not be liquefied. Attempts were made by J. O. Natterer (1844–55) to liquefy these gases by the use of high pressures; he devised a pump for the purpose capable of producing pressures exceeding 3000 atm., but his efforts met with failure. The general opinion in the middle of last century was, therefore, that certain gases could not be liquefied in any circumstances, and these were called **permanent gases**; this conclusion was shortly to be proved incorrect.

The Critical State.—The essential condition for the liquefaction of a gas was discovered by T. Andrews (1869) as the result of a study of the temperature, pressure and volume relationships of carbon dioxide. The apparatus employed is depicted in Fig. 90; it consisted of two similar graduated glass tubes, one containing carbon dioxide and the other air, each enclosed by a drop of mercury. These were fixed by means of

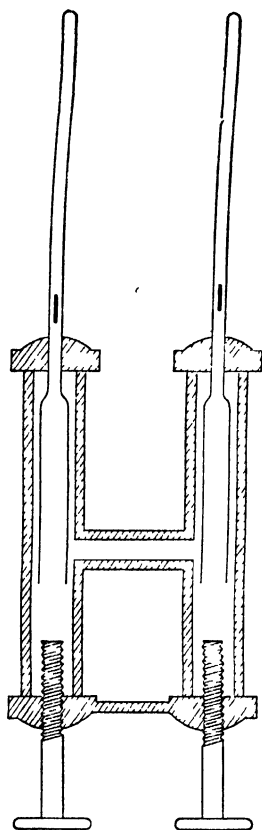


Fig 90. Compression of gases (Andrews)

liquid-tight joints into metal compression chambers, connected together and containing water to which pressure could be applied by means of screw plungers. The upper portions of the tubes were kept at a definite temperature, and the volume of carbon dioxide at different temperatures and pressures was measured; the value of the pressure was calculated from the volume of air. The volumes of carbon dioxide, at different pressures and various constant temperatures, were plotted by Andrews in the form of a series of isothermal curves as shown in Fig. 91.

At the lowest temperature employed, 13.1° c., the carbon dioxide is entirely gaseous at low pressures, as at *A*; on increasing the pressure the volume decreases, as indicated by the curve *AB*, approximately in accordance with Boyle's law. At the pressure *B*, however, liquefaction commences and the volume decreases rapidly as the gas is converted into liquid with a much higher density; at *C* the carbon dioxide has been completely liquefied, and the steep curve *CD* represents the fact that the liquid is not easily compressed and undergoes only a small change of volume as a result of increase of pressure. It should be noted that the portion *AB* of the curve represents gas only, and *CD* liquid only; along *BC*, however, gas and liquid can coexist, but since *BC* is parallel to the volume axis it follows that the pressure remains constant while these two phases are present irrespective of their relative amounts. The importance of this

pressure at which a liquid and its vapor are in equilibrium at a given temperature, that is the **vapor pressure** of the liquid, will be considered later. The pressure-volume curve at 21.5° c. is similar to that at the lower temperature, except that the horizontal portion, over which liquefaction occurs, is shorter; in fact as the temperature is raised this section becomes less and less, as indicated by the dotted "boundary curve," and finally at 31.1° c. it is merely a point. Above 31.1° there is no indication whatever of liquefaction, and Andrews found that if this

temperature was exceeded carbon dioxide could not be liquefied even at pressures of 300 or 400 atm., whereas below 31.1° a pressure of 75 atm. was sufficient. It is evident, therefore, that for carbon dioxide there is a limit of temperature above which the gas cannot be liquefied no matter what the pressure. Further investigation, since the time of Andrews, has shown that all other gases exhibit the same property, the temperature depending on the nature of the gas. The reason for the failure to liquefy the so-called permanent gases is now clear: they had not been cooled sufficiently before applying pressure.

↷ The maximum temperature at which a gas can be liquefied, that is the temperature above which liquid cannot exist, is called the **critical temperature**, and the pressure required to cause liquefaction at this temperature is the **critical pressure**; the latter is, incidentally, the highest vapor pressure the liquid can exert.↷ The **critical volume** is the volume occupied by 1 mole of the substance * at the critical temperature and pressure, and the pressure-volume curve for the critical temperature is referred to as the **critical isotherm**. The point *P* in Fig. 91 represents carbon dioxide in its critical state, the temperature, pressure and volume being the critical values. It may be remarked that the term **vapor** is used to describe a gaseous substance when its temperature is below the critical value; a vapor can, therefore, be condensed to a liquid by pressure alone.†

Measurement of Critical Constants.—The direct method, as used by Andrews, for the measurement of the critical constants of a substance is not always convenient, and various other procedures have been adopted. It was observed by Cagniard de la Tour (1822) that when a liquid, e.g., ether or carbon disulfide, is heated in closed space, the surface of separation between liquid and vapor disappears at a definite temperature; on cooling the two layers re-appear at the same temperature. In the course of his preliminary work, T. Andrews (1863) noted that these phenomena occurred at the critical point; he stated that at this temperature "the surface of demarcation between liquid and gas became fainter and at last disappeared. The space was then occupied by a homogeneous fluid which

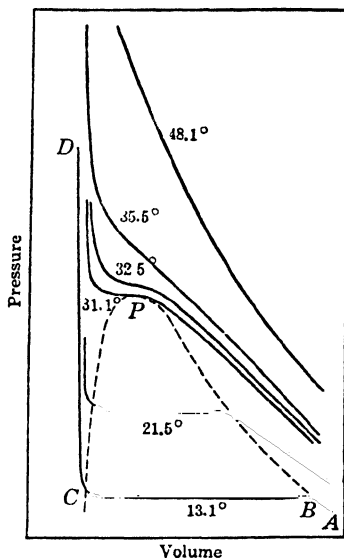


FIG. 91. Pressure-volume curves of carbon dioxide

* The values quoted in the literature are generally in terms of the volume of the gas at s.t.p., i.e., as multiples of 22.414 liters.

† It is claimed that the general idea of the critical state was arrived at independently by D. I. Mendeléeff in 1861.

exhibited," when the temperature was lowered slightly, "a peculiar appearance of moving or flickering striae," and the surface of separation between liquid and vapor was again evident.

These observations have been utilized in methods for determining critical temperatures and pressures. A liquid, e.g., water or benzene, is heated gently in a sealed tube and the temperature noted at which the meniscus disappears, and at which it re-appears on cooling; this is the critical temperature. If the

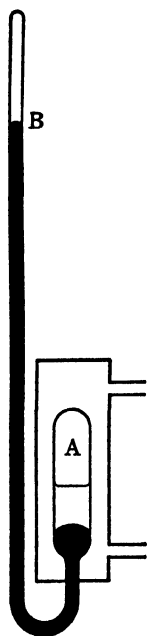


FIG. 92. Observation of critical phenomena

experiment is made in an apparatus similar to that used by C. de la Tour (Fig. 92), in which the liquid and vapor *only* are enclosed by mercury in the bulb *A* surrounded by a heating jacket, both critical temperature and pressure can be determined. The bulb is connected to a manometer *B* containing air, and the pressure is read at the temperature at which the surface between liquid and vapor disappears. The critical pressure alone can be determined by the method of M. Altschul (1893). The liquid is placed in a narrow sloping tube sealed at its upper end, and enclosed at the lower end by a drop of mercury, connected through a hydrocarbon oil to a compression pump. The upper end of the tube is kept above the critical temperature whereas the lower end is below it; at some intermediate position the tube is just at the critical temperature. The pressure on the liquid is gradually increased by means of the pump and the surface of separation between liquid and gas moves up the tube; when it reaches the point at which the temperature is the critical value, the meniscus disappears and the critical pressure is then indicated by a pressure gauge attached to the pump.

It was noted above that when liquid and vapor coexist the pressure, at a given temperature, is independent of their relative volumes; above the critical temperature, however, when only gas can exist, the pressure must vary with the volume. The method of determining critical constants described by L. Cailletet and E. Colardeau (1888) depends on these facts. Different volumes of liquid are placed in a steel tube and the pressures, at constant volume, are measured at a series of increasing temperatures, and the results plotted. As long as the pressure

is independent of the volume the temperature is below the critical value, but beyond this point the pressure increases with the amount of substance used; both critical pressure and temperature can thus be determined.

In the ingenious method of P. A. Bond and D. A. Williams (1931), a strong tube, approximately half filled with the experimental liquid, is supported almost horizontally on a central knife edge; a wire attached to the lower end of the tube is fixed to the arm of a balance placed vertically above it. The tube is heated in an air jacket above the critical point so as to permit complete vaporization, and after allowing time for a uniform distribution of vapor to be attained, the balance is suitably adjusted. The system is now allowed to cool; when the critical temperature is reached, liquid forms, runs to one end of the tube and so causes a disturbance of the balance.¹

The Law of the Rectilinear Diameter: Critical Volumes.—The critical volume is best obtained with the aid of the rule known as the **law of the rectilinear diameter**, discovered by L. Cailletet and E. Mathias (1886) and subsequently verified by S. Young (1900) and others. The law states that the mean of the densities of any substance in the state of liquid and of saturated vapor,* at the same temperature, is a linear function of the temperature. The densities of the liquid and of saturated vapor in equilibrium with it, are known as the **orthobaric densities**, and if ρ_t is the arithmetical mean at any temperature t , then

$$\rho_t = \rho_0 + \alpha t, \quad (1)$$

where ρ_0 and α are constant for the given substance. The applicability of the rule to *n*-pentane is shown in Fig. 93; the points on *AC* give the

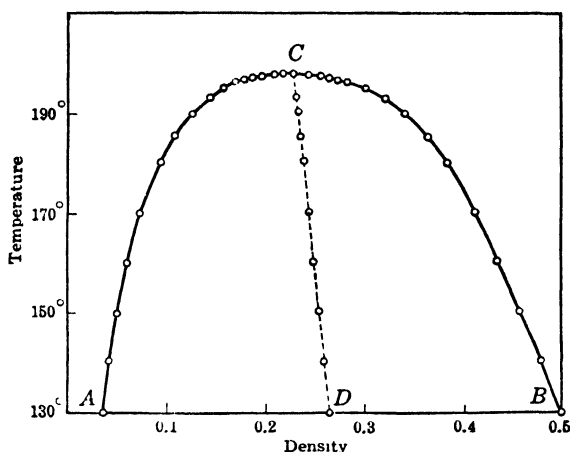


FIG. 93. Orthobaric densities and rectilinear diameter of *n*-pentane

densities of the saturated vapor and on *BC* of the liquid at various temperatures, *C* being the critical point. The mean densities are plotted along *CD*, and the close approximation to a straight line is evident; good agreement is found with the linear equation

$$\rho_t = 0.3231 - 0.00046t. \quad (2)$$

To determine the critical volume of the substance, therefore, the orthobaric densities are measured at a number of temperatures near the critical point; the mean values are then either plotted against temperature, or a linear equation of the type just given is deduced, and from a knowledge of the critical temperature the mean density at this point can be obtained by extrapolation or calculation. It is seen from Fig. 93 that at the

* These states correspond to the points on the two "arms" of the dotted boundary curve of the Andrews diagram (Fig. 91).

critical temperature the mean density is equal to the density of both the liquid and vapor, which are in fact identical; from this value the critical volume can be immediately calculated.

The determination of orthobaric densities is made by enclosing a known mass m of liquid in a graduated sealed tube and heating to a known temperature; the volumes of liquid v_L and of saturated vapor v_V are read off from the graduations, and if ρ_L and ρ_V are the corresponding densities, it follows that

$$m = \rho_L v_L + \rho_V v_V. \quad (3)$$

The experiment is repeated at the same temperature with a different mass of substance, and from the two equations ρ_L and ρ_V can be evaluated.²

The critical data for a number of common gases are given in Table 64 (see also Table 69, p. 456).

TABLE 64. CRITICAL DATA

Sub- stance	Temp.	Press.	Dens.	Sub- stance	Temp.	Press.	Dens.
He	5.2° K.	2.26 atm.	0.066 g./cc.	CO ₂	304.2° K.	72.9 atm.	0.460 g./cc.
H ₂	33.2°	12.8	0.031	N ₂ O	309.6°	71.9	0.454
N ₂	126.0°	33.5	0.311	HCl	324.5°	81.6	0.406
CO	134.4°	34.6	0.311	H ₂ S	373.5°	89.0	0.268
A	150.7°	48.0	0.510	NH ₃	405.5°	111.5	0.236
O ₂	154.3°	49.7	0.430	Cl ₂	417.1°	76.1	0.573
CH ₄	190.2°	45.6	0.162	SO ₂	430.3°	77.6	0.513

The law of Cailletet and Mathias is in harmony with the concept of "holes" in liquids which has been employed in connection with some theories of viscosity and the liquid state (see p. 502). Holes may be regarded as playing the same part in the liquid phase as molecules do in the gas phase, the "holes" moving about in the liquid just as do the molecules in a gas. At the critical temperature there will be exactly as much matter in unit volume of otherwise empty space as there are holes in unit volume of matter, since at this point liquid and vapor become identical. As the temperature is lowered the number of holes in the liquid decreases but so also does the number of molecules in the vapor to an equivalent extent; the mean density of the two phases might thus be expected to remain constant. In view of the normal contraction which matter undergoes on cooling, however, there should be actually a small linear increase in the mean density as the temperature is lowered, as required by the law of the rectilinear diameter (see Fig. 93).

Continuity of State.—The transition from liquid to vapor, or the reverse, generally occurs suddenly, the two states being sharply distinguished from each other, as for example by the portions *AB* and *CD* of the 13.1° C. isothermal in Fig. 91. It is evident, however, that as the critical temperature is approached the volumes, and hence the densities, of liquid and of saturated vapor become closer, and are identical at the critical point. This implies that there is no difference between liquid and gas at this temperature, a conclusion in agreement with the observation

that the surface of separation between the two phases also disappears at this point. The possibility arises, therefore, of a gradual transition from one phase to the other, for it is obvious that just below the critical temperature there can be little difference between liquid and vapor.

The idea of a continuity from the liquid to the gaseous states can be brought out in other ways. If a gas at *A* (Fig. 94) is compressed at a constant temperature below the critical point, the change of volume is represented by a curve of the type *ABCD*, there being a discontinuous change from gas to liquid along *BC* where the two phases coexist. Suppose, however, the volume were kept constant and the temperature raised from *A* to *L*, the pressure increasing at the same time; the pressure is now held constant while the temperature, and consequently the volume, is decreased to *D*. The system which was entirely gaseous at *A* is now completely liquid at *D* which is on the same isotherm; the change has

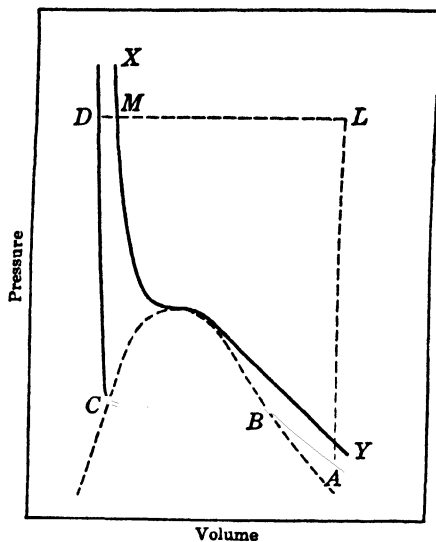


Fig. 94. Continuity of state

been made from gas to liquid, but there has been no sharp discontinuity to indicate a sudden transition from one state to the other, and at no time has there been more than one phase in existence. The change from gas to liquid actually occurs in the process represented by LD ; to the right of the critical isothermal curve XY the system is gaseous, but to the left it is liquid, and consequently the transition must take place at M . As the temperature is decreased from L to D the density of the gas increases, and presumably at M the molecules are close enough for the forces of cohesion to lead to the formation of a definite bounding surface or meniscus. Nevertheless, the change from gas to liquid is gradual, and it may be regarded as a continuous transition rather than as a discontinuous one.

If a vapor is compressed in a perfectly clean and smooth vessel, it is often found that the pressure at which liquefaction should normally occur can be exceeded without condensation taking place; this may be indicated diagrammatically by the dotted line BE in Fig. 95, where $ABCD$ is the normal isotherm. Any attempt to increase the pressure beyond E results in sudden liquefaction. Similarly, if the pressure of a pure liquid, free from dissolved air, is reduced, it is possible for the unstable curve CG to be realized, although there will be an increasing tendency for sudden, almost explosive, evaporation to occur; the behavior is here identical

with the familiar phenomenon of "superheating." The suggestion was made by J. Thomson (1871) that the ideal behavior of a gas should be

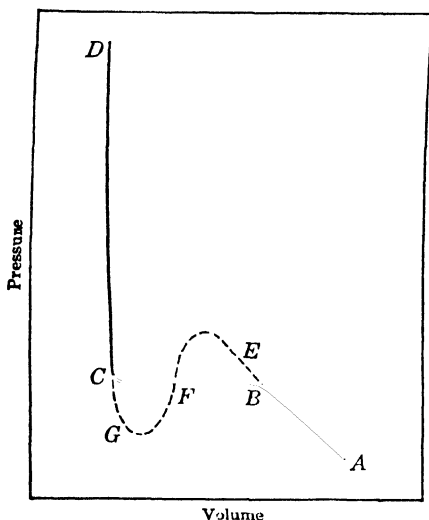


FIG. 95. Continuity of state

represented by a continuous curve *EFG* connecting *BE* and *CG*, thus emphasizing from another standpoint the essential continuity of gaseous and liquid states. Although the conditions between *E* and *G* are not realizable in practice, Thomson thought they might have a theoretical significance. This significance lies in the fact that whereas it is difficult to express the discontinuous behavior, as represented by *ABCD*, in the form of an equation, the continuous curve *ABEFGCD* may in fact be given simple mathematical expression.³

Indefiniteness of the Critical State.—As far back as 1880 various workers made observations which suggested that the liquid state still persisted above the apparent critical temperature, and the experiments of I. Traube and his collaborators (1904) indicated that even after the meniscus had disappeared, the density of the system was not uniform. Various attempts were made to explain away these observations as being due to experimental error or other causes, but repetition of the work by O. Maass (1933 *et seq.*), and others, appears to establish the correctness of the result. Although when a liquid is heated in a closed space the meniscus disappears at a definite temperature, differences of density persist even if the system is stirred; a few degrees above this temperature, however, the density becomes uniform at the true critical point. By taking into account the mutual interactions between molecules, which are supposed to form clusters of varying size, it has been shown theoretically (J. Mayer, 1937 *et seq.*) that a "critical region," between the temperature at which the meniscus disappears and that at which liquid can no longer exist, is to be expected. For the present, however, it will be assumed that each substance has a definite critical point above which liquid cannot exist; this is undoubtedly true, although it may not coincide with the temperature at which the meniscus disappears.⁴

Van der Waals's Equation and the Critical State.—The van der Waals equation (p. 289)

$$\left(P + \frac{a}{V^2}\right)(V - b) = RT \quad (4)$$

can be readily transformed into the cubic equation

$$V^3 - \left(b + \frac{RT}{P}\right)V^2 + \frac{a}{P}V - \frac{ab}{P} = 0 \quad (5)$$

and consequently is of the form required for Thomson's theoretical gas-liquid curve. Below a certain value of the temperature T the equation can have three real roots,* but as T is increased these approach one another, and eventually become identical. The \sim -shaped portion of the pressure-volume curve thus becomes smaller and smaller until it is reduced to a point of horizontal inflection. For still higher values of T , two of the three roots of (5) are imaginary and only one is real, so that the curve will show no maximum, minimum or inflection. These purely mathematical deductions from (5) are equivalent to the facts that the horizontal portion of the isothermal curve in Fig. 91 becomes shorter as the temperature is raised, that it is merely a point at the critical temperature, and that above this temperature there is a complete absence of discontinuity in the pressure-volume curve. The van der Waals curve having a point of horizontal inflection should, therefore, be obtained at the critical temperature.

The criterion of a point of this type is that both first and second derivatives should be zero, and consequently it may be assumed, on the basis of the above arguments, that $(\partial P/\partial V)_T$ and $(\partial^2 P/\partial V^2)_T$ are both zero at the critical temperature. The van der Waals equation may be written in the form

$$P = \frac{RT}{V-b} - \frac{a}{V^2}, \quad (6)$$

$$\therefore \left(\frac{\partial P}{\partial V}\right)_T = -\frac{RT}{(V-b)^2} + \frac{2a}{V^3} \quad \text{and} \quad \left(\frac{\partial^2 P}{\partial V^2}\right)_T = \frac{2RT}{(V-b)^3} - \frac{6a}{V^4}, \quad (7)$$

and hence for the critical isotherm,

$$-\frac{RT_c}{(V_c-b)^2} + \frac{2a}{V_c^3} = 0 \quad (8)$$

and

$$\frac{2RT_c}{(V_c-b)^3} - \frac{6a}{V_c^4} = 0, \quad (9)$$

the suffix c being used to indicate the critical state. Dividing (9) by (8), it is seen that $V_c = 3b$, and substitution in either gives $T_c = 8a/27Rb$; if these values are inserted in (6) the result $P_c = a/27b^2$ is obtained; hence,

$$(i) \quad V_c = 3b, \quad (ii) \quad T_c = \frac{8a}{27Rb}, \quad (iii) \quad P_c = \frac{a}{27b^2}. \quad (10)$$

The applicability of the van der Waals equation may, therefore, be tested by calculating the critical constants of a gas for which a and b have been determined from pressure-volume, or other, measurements (see p. 290). As already mentioned (p. 295) these quantities are not strictly constant,

* The three roots correspond to the points B , F and C in Fig. 95; B and C represent saturated vapor and liquid respectively, but F has no physical significance.

and so the test can only be regarded as approximate. By means of the following values of a , b and R ,

$$a = 4.37 \text{ liter}^2\text{-atm.}, \quad b = 0.0515 \text{ liters}, \quad R = 0.082 \text{ liter-atm. deg.}^{-1},$$

the behavior of 1 mole of carbon dioxide at ordinary temperatures may be represented with fair accuracy. It may consequently be deduced by the relationships given above that

$$V_c = 0.154 \text{ liters per mole}; \quad T_c = 306^\circ \text{ K.}; \quad P_c = 61 \text{ atm.},$$

which may be compared with the observed values of 0.10 liters per mole, 304° K. , and 73 atm. respectively. The agreement is reasonably good, but it would have been less so if the value of a at higher temperatures, e.g., 3.41 at 100° C. , had been employed; this method of testing the van der Waals equation is, therefore, not satisfactory.

From the equations for P_c , V_c and T_c in (10) the van der Waals constants can be derived; the results are

$$a = \frac{27R^2T_c}{64p_c} \quad \text{and} \quad b = \frac{RT_c}{8p_c}, \quad (11)$$

and so an alternative test is possible by calculating a and b from the known critical temperature and pressure of a particular substance and then finding how far the resulting van der Waals equation represents the pressure-volume relationships of the gas. For example, R. E. Baynes (1880) determined a and b for ethylene from its critical data, and obtained the values 0.00786 and 0.00224 respectively, in Amagat units (p. 289); inserting these figures in the van der Waals equation he calculated PV for various pressures at 20° C. The results obtained in this manner, assuming PV to be 1.000 at 1 atm. pressure, are compared with the experimental data on the compressibility of ethylene in Table 65. The agree-

TABLE 65. OBSERVED AND CALCULATED PRESSURE-VOLUME PRODUCT OF ETHYLENE AT 20° C.

PV			PV		
$P \text{ atm.}$	obs.	calc.	$P \text{ atm.}$	obs.	calc.
31.6	0.914	0.895	133.3	0.520	0.520
45.8	0.781	0.782	176.0	0.643	0.642
72.9	0.416	0.387	233.6	0.807	0.805
82.2	0.399	0.392	282.2	0.941	0.940
94.5	0.413	0.413	398.7	1.248	1.254

ment is certainly good, but it may be to some extent due to the fact that the experimental temperature was only 10° from the critical value; at higher or lower temperatures the discrepancies would probably be more marked. It was mentioned on p. 290 that the a and b values generally quoted in the literature (e.g., Table 64) are based on critical data, use being made of the relationships of (11); the results are, therefore, strictly applicable only in the vicinity of the critical temperature.⁵

According to (10, i) the critical volume should be equal to $3b$, that is, $V_c/b = 3$; the necessary data for verifying this relationship are limited, but the following experimental results are of interest.

	Argon	Hydrogen	Carbon dioxide	Air
V_c/b	1.41	2.80	1.86	1.89

The ratio of V_c to b is evidently not constant, and it is generally nearer to 2 than the value, viz., 3, required by the van der Waals equation. An attempt was made by D. Berthelot (1900) to evaluate V_c/b by assuming that b would be equal to the limiting volume V_0 of the liquid at 0°K . Using the rule of Cailletet and Mathias the mean density of a substance at 0°K . can be determined by extrapolation, and since the density of the vapor in equilibrium with the liquid would then be zero, the limiting density of the liquid should be double the mean density. From this the limiting zero volume was calculated, and it was found that V_c/V_0 was very nearly 4 for a number of substances, e.g., oxygen, chlorine, carbon dioxide, sulfur dioxide, ethylene, carbon tetrachloride and benzene. Berthelot concluded, therefore, that $V_c/b = 4$ (see p. 295); this involves the assumption that V_0 and b are identical, which should be the case according to the van der Waals equation. There is obviously a discrepancy between the two methods of evaluating V_c/b which has not yet been cleared up.

A further consequence of the van der Waals equation can be deduced by combining the three parts of (10); thus,

$$RT_c/P_cV_c = 8/3 = 2.67, \quad (12)$$

so that the critical fraction should be constant irrespective of the nature of the substance. For 1 mole of an ideal gas the ratio RT/PV should always be unity, and so at the critical temperature the product PV of any gas obeying the van der Waals equation should be $1/2.67$ of the value for an ideal gas. The data in Table 66 show, however, that although

TABLE 66. RATIO OF IDEAL TO OBSERVED PV AT CRITICAL TEMPERATURE

Substance	RT_c/P_cV_c	Substance	RT_c/P_cV_c	Substance	RT_c/P_cV_c
Hydrogen	3.05	Nitrous oxide	3.62	<i>n</i> -Pentane	3.76
Helium	3.08	Carbon dioxide	3.64	Ethyl ether	3.81
Nitrogen	3.42	Carbon tetrachloride	3.68	Ethyl acetate	3.88
Oxygen	3.42	Ethylene	3.64	Chlorobenzene	3.81
Argon	3.35	Benzene	3.71	Stannic chloride	3.75

RT_c/P_cV_c is almost constant for a variety of elements and compounds, it is appreciably greater than 2.67. Substances containing hydroxyl groups, which are definitely known to be associated in the liquid state, give much higher values for the RT_c/P_cV_c ratio; thus for water it is 4.39, for ethyl alcohol 4.02, and for acetic acid 4.99.

It was shown on p. 291 that the Boyle temperature, where Boyle's law is obeyed over a range of low pressures, is given by the expression

$T_B = a/Rb$, and since $T_c = 8a/27Rb$, it follows that

$$T_B = 27/8T_c = 3.375T_c. \quad (13)$$

The Boyle temperature of any gas should thus be 3.375 times the critical temperature; the experimental data show that although T_B/T_c is approximately constant for several gases, the value is closer to 2.5. This low result is in general harmony with the fact that RT_c/P_cV_c is larger than expected theoretically.

Other Equations of State.—In reviewing the results of the application of van der Waals's equation to the critical state it is apparent that although it leads to conclusions which are qualitatively correct, they are not satisfactory from the quantitative standpoint. It is of interest, therefore, to examine other equations of state to see if they give better agreement with experiment. The Clausius equation (p. 295), like that of van der Waals, is of the third order with respect to V , and gives pressure-volume curves of the same type; at the critical point, therefore, the first and second derivatives with respect to the volume, i.e., $(\partial P/\partial V)_T$ and $(\partial^2 P/\partial V^2)_T$, may be equated to zero, when it follows that

$$(i) \quad V_c = 3b + 2c, \quad (ii) \quad T_c^2 = \frac{8a}{27R(b+c)}, \quad (iii) \quad P_c = \frac{RT_c}{8(b+c)}, \quad (14)$$

and
$$\frac{RT_c}{P_cV_c} = \frac{8(b+c)}{3b+2c}. \quad (15)$$

The Clausius equation involves four constants, a , b , c and R , but for reasons which will appear shortly it is necessary that any equation of state should have no more than three *independent* constants; the quantity c may, therefore, be put equal to kb , where k is a universal constant, which may be evaluated in the following manner. By the use of a method analogous to that described on p. 291, the Boyle temperature, according to the Clausius equation, should be given by

$$T_B^2 = \frac{a}{Rb}, \quad (16)$$

and from (14, ii),

$$T_B^2 = 27(1+k)T_c^2/8. \quad (17)$$

If T_B/T_c is assumed to have the experimental value of 2.5, it follows that

$$k = 0.85,$$

and using this, in conjunction with $c = kb$, in (15), there results

$$RT_c/P_cV_c = 3.15, \quad (18)$$

which is a slight improvement on the result derived from the van der Waals equation. On the whole the Clausius equation does not represent a great advance, and since it suffers from the disadvantage of involving four constants, it has not come into general use.

The Dieterici equation (p. 296) has only three independent constants, and also satisfies the requirements of having three real roots, provided T is below a certain value; the roots eventually become equal to each other, when the pressure-volume curve corresponds to the critical isothermal and has an inflec-

tion at the critical point. By equating the first and second derivatives to zero in the usual manner, it is found that

$$(i) \quad V_c = 2b, \quad (ii) \quad T_c = \frac{a}{4Rb}, \quad (iii) \quad P_c = \frac{a}{4e^2b^2} = \frac{a}{29.56b^2}, \quad (19)$$

and

$$RT_c/P_cV_c = \frac{1}{2}e^2 = 3.695, \quad (20)$$

where e is the base of natural logarithms. As far as V_c and RT_c/P_cV_c are concerned the agreement with experiment is very satisfactory; the ratio of the Boyle temperature to the critical temperature, according to the Dieterici equation, that is

$$T_B = a/Rb = 4T_c, \quad (21)$$

is, however, too large.

In spite of its limitations the van der Waals equation is still widely employed as a basis for a consideration of the behavior of gases, and also of liquids. It is true that the Dieterici equation is superior in some respects, but as already recorded (p. 296) it is to be regarded as involving substantially the same ideas; as it is somewhat more difficult to handle mathematically it has not achieved the popularity of the van der Waals relationship. It should be noted, too, that the a and b values in both equations vary with temperature, and so it is not possible to write either in a form applicable over a range of temperature. The two relationships are thus inevitably approximate, and so that van der Waals equation, which is the simpler and the more readily understood, is generally used. It is important, however, that its limitations should be borne in mind.⁶

Law of Corresponding States.—One of the most striking consequences of his equation was pointed out by J. D. van der Waals in 1881. If the pressure, volume and temperature of a gas are expressed in terms of the critical pressure, volume and temperature, respectively, thus,

$$P = \pi P_c, \quad V = \phi V_c, \quad T = \theta T_c, \quad (22)$$

the van der Waals equation (4) becomes

$$\left(\pi P_c + \frac{a}{\phi^2 V_c^2} \right) (\phi V_c - b) = R\theta T_c, \quad (23)$$

and if the values for P_c , V_c and T_c , as given by (10), are introduced, it follows that

$$\begin{aligned} \left(\frac{\pi a}{27b^2} + \frac{a}{9\phi^2 b^2} \right) (3\phi b - b) &= \frac{8a\theta}{27b}; \\ \therefore \left(\pi + \frac{3}{\phi^2} \right) (3\phi - 1) &= 8\theta. \end{aligned} \quad (24)$$

The quantities π , ϕ and θ are called the **reduced pressure, volume and temperature**, respectively, and (24) is the **reduced equation of state**. The important point about this equation is that it is perfectly general and does not involve the constants a , b and R , and should be applicable to all substances. Further, since it may be concluded from the discussion given above that the van der Waals equation can represent, at least

qualitatively, the behavior of a substance in both *liquid and gaseous* phases, the reduced equation should likewise hold for each of these forms. Since it is of the third order with respect to ϕ there will be three solutions to the equation below the critical temperature, one corresponding to liquid, another to vapor, whereas the intermediate one has no physical significance (see p. 435, footnote).

An examination of (24) shows that if any two or more substances have the same reduced pressures π , that is to say, their pressures are the same fraction or multiple π of their respective critical pressures, and are at equal reduced temperatures θ , their temperatures being the same fraction or multiple θ of their respective critical temperatures, then their reduced volumes ϕ should be equal, i.e., V/V_c should be the same for all. The substances are then said to be in **corresponding states**, and the above statement may be taken as an expression of the **law of corresponding states**. The accuracy of the law is shown by the data in Table 67 taken from

TABLE 67. LAW OF CORRESPONDING STATES

$$\pi = 0.08846 \quad \theta = 0.73-0.75$$

Substance	ϕ liquid	ϕ vapor	Substance	ϕ liquid	ϕ vapor
Benzene	0.4065	28.3	Methyl formate	0.4001	29.3
<i>n</i> -Pentane	0.4061	28.4	Carbon tetrachloride	0.4078	27.5
<i>n</i> -Octane	0.4006	29.3	Stannic chloride	0.4031	27.2
Ethyl ether	0.4030	28.3	Fluorobenzene	0.4067	28.4

those collected by S. Young (1908); the reduced pressure π is 0.08846 in every case, but the reduced temperature θ varies from 0.73 to 0.75. The observed reduced orthobaric volumes are given for the liquids and saturated vapors in equilibrium. It appears from these and other results that the general form of the law of corresponding states is correct, especially for nonassociated substances of similar constitution, but it must be emphasized that this does not necessarily provide support for the van der Waals equation. If the data in Table 67 are inserted in (24) considerable discrepancies will be found, especially for the saturated vapor, and the work of C. Raveau (1896), and of others, has shown that the reduced equation (24) suffers from the same limitations as does the original van der Waals equation from which it was derived.

The important fact to remember is that any equation of state involving only two arbitrary constants in addition to R , e.g., the Dieterici equation, can be converted into an equation involving only the reduced quantities π , ϕ and θ ; the generalized form of the law of corresponding states then follows directly. Since this law appears to hold for a number of substances, it is usually accepted that an equation of state should involve not more than three independent constants.⁷

Production of Low Temperatures.—As a consequence of Andrews's discovery that a gas had to be cooled below a certain critical temperature before it

could be liquefied, attention was concentrated on the problem of obtaining low temperatures. In 1877 R. P. Pictet liquefied oxygen, and L. Cailletet independently liquefied this gas, as well as carbon monoxide and nitrogen. The latter worker used the principle of adiabatic expansion, described below, whereas the former adopted a "cascade" method. By evaporation of liquid sulfur dioxide under reduced pressure a temperature of -65°C . was obtained, and cooled by this means carbon dioxide was liquefied in quantity. When this liquid was boiled at low pressure the temperature fell to about -130°C ., which was sufficient to bring about liquefaction of compressed oxygen. By utilizing liquid oxygen in the cascade, S. von Wroblewski (1884) obtained indications of liquid hydrogen but appreciable quantities of this liquid did not become available until J. Dewar (1900) employed the Joule-Thomson effect (p. 286) after preliminary cooling to -200°C . by the evaporation of liquid nitrogen. The most difficult gas to condense proved to be helium, but this was liquefied by H. Kamerlingh Onnes (1908) who also made use of the Joule-Thomson effect after cooling the gas to about -255°C . with liquid hydrogen.

The methods at present in use for liquefying gases involve either adiabatic expansion or the Joule-Thomson effect, and these will be described briefly. When a gas expands and does work against the external pressure, its temperature must fall, provided the process is carried out in a heat-insulated system, that is, adiabatically. It has been shown on p. 199 that if an ideal gas undergoes an adiabatic change from pressure P_1 and volume V_1 to pressure P_2 and volume V_2 , then $P_1 V_1^{\gamma} = P_2 V_2^{\gamma}$, where γ is the ratio of the specific heats of the gas. If the initial and final temperatures are T_1 and T_2 , then $P_1 V_1 = RT_1$ and $P_2 V_2 = RT_2$, and if V_1 and V_2 are eliminated from these three equations, it is seen that

$$(T_1/T_2)^{\gamma} = (P_1/P_2)^{\gamma-1}. \quad (25)$$

The fall of temperature in an adiabatic expansion is thus greater the larger the difference between initial and final pressures.

In the method for the commercial liquefaction of air, devised by G. Claude (1900-05), the gas is compressed to about 40 atm. and then allowed to expand, sometimes in two stages, to 1 atm. in an expansion engine where it does external work, which is frequently utilized for driving a dynamo. The emerging gas has been cooled appreciably, and is used to lower the temperature of the incoming compressed gas before it enters the engine. When the process is running properly the gas at 40 atm. is first cooled to -80°C . and divided into two parts; one portion passes through the expansion engine, where its temperature and pressure are both decreased, and the other, which retains its high pressure, is cooled sufficiently by means of the gas leaving the engine for liquefaction to occur. In modified forms of the Claude process pressures up to 200 atm. are used.

One of the chief difficulties experienced by Claude was to obtain efficient lubrication of the piston of the expansion engine at low temperatures; this problem is more serious the lower the temperature. In the engine designed by P. Kapitza (1934) for the liquefaction of helium, there is a small gap between the piston and the cylinder so that there is no actual contact and lubrication is unnecessary; loss of gas is minimized by the very rapid motion of the piston. With this apparatus the emerging gas can be cooled down to -263°C . In the adiabatic expansion method of F. Simon (1932) a small metal cylinder, cooled by liquid hydrogen, is filled with compressed helium; the pressure is

gradually released and work is done in forcing out some of the gas, with the result that the temperature of the remainder falls sufficiently for liquefaction to take place.

Cooling by the Joule-Thomson effect has been employed for large-scale work by W. Hampson (1895) and by C. von Linde (1895). The gas at a pressure of about 200 atm. is allowed to stream through a valve or throttle, as a result of which its pressure is reduced to about 40 atm. (Linde) or 1 atm. (Hampson). The external work done in this expansion, i.e., $P_2V_2 - P_1V_1$ (cf. p. 132), is negligible, but considerable *internal* work is done in overcoming the molecular attraction, thus giving Joule-Thomson cooling. As pointed out on p. 294 there is a definite inversion temperature at each pressure above which a gas becomes warmer on streaming through a throttle, whereas below it there is a fall of temperature on expansion. Operation above the inversion point for the given pressure should, therefore, be avoided. It can be shown theoretically that the greatest efficiency is obtained when the initial temperature and pressure have the inversion values, and an attempt is made to approximate to these conditions in practice. In the actual liquefier the gas after passing through the throttle is used to cool the incoming gas and eventually liquefaction occurs. The Joule-Thomson effect is applied for the purpose of liquefying appreciable quantities of hydrogen and helium.

Two other methods of obtaining low temperatures may be mentioned. When a gas is adsorbed on charcoal heat is evolved (see p. 1195); consequently, if the adsorbed gas is pumped off the temperature must fall. By making use of the principle F. Simon (1926) liquefied helium on a small scale: considerable quantities of the gas are adsorbed by activated charcoal cooled by liquid hydrogen. The surrounding space is then evacuated so as to minimize heat conduction, and the helium is pumped off; the temperature of the charcoal falls below the critical point of helium, and if the compressed gas is passed through a tube cooled in this manner liquefaction occurs.

Adiabatic Demagnetization.—The suggestion for the production of low temperatures by adiabatic demagnetization was made independently by P. Debye (1926) and W. F. Giauque (1926). The entropy of a paramagnetic solid may be regarded as partly thermal and partly associated with its magnetic properties; if the material is placed in a magnetic field there is a tendency for the atomic magnetic dipoles (cf. p. 615) to arrange themselves in a definite manner. Since entropy is a measure of the randomness of distribution (p. 229), application of the magnetic field causes the entropy of the system to decrease. If the substance is now thermally insulated, so that the magnetic field can be removed adiabatically, the total entropy of the system must remain constant, because in an adiabatic process there is no change of entropy. The magnetic atomic dipoles return to a random arrangement, however, when the field is removed; hence adiabatic demagnetization is accompanied by an increase of magnetic entropy, and consequently there must be a decrease of thermal entropy. The latter is determined essentially by thermal movements and a decrease must result in a fall of temperature. The magnetic entropy is only appreciable, in comparison with the thermal portion, at low temperatures, and so the method of cooling by adiabatic demagnetization is most effective under these conditions. The paramagnetic solid is first cooled down to about 1° K. by the evaporation of liquid helium, a small amount of the gas being employed to maintain heat contact between the solid and the liquid helium. The magnetic field, as high as 30,000 gauss if possible, is then switched on; the temperature

risers, but the heat is conducted away by the surrounding gas. When the temperature has again fallen to that of the liquid helium the conducting gas is pumped off, so that the vessel containing the paramagnetic solid is thermally insulated; the removal of the magnetic field is now accompanied by a fall of temperature. Rare-earth salts, e.g., gadolinium, cerium- and dysprosium compounds, were employed in the early studies of the magnetic method of cooling, but later it was found that good results could be obtained with such substances as ferric ammonium alum (F. Simon and N. K rti, 1935) and potassium chrome alum (W. J. de Haas and E. C. Wiersma, 1935). The lowest temperature recorded, 0.003°K. , was obtained with the latter substance, and temperatures of about 0.01°K. to 0.02°K. can now be produced without serious difficulty by the magnetic method. The possibility of obtaining such low temperatures has opened up a new and highly important field for the investigation of properties, such as electrical conductivity, which are abnormal in these regions.⁸

VAPOR PRESSURE AND VAPORIZATION

Vapor Pressure.—Earlier in this chapter the subject of gas-liquid equilibria was discussed mainly from the standpoint of the gas; it will now be considered with special reference to the liquid. It has been seen that if a gas is compressed below its critical temperature, liquefaction commences at a certain pressure which remains constant, at each temperature, as long as liquid and vapor are present. This pressure, under which vapor and liquid can co-exist in equilibrium, is the **saturation vapor pressure** or, briefly, the **vapor pressure**, of the liquid at the given temperature. It is seen from Fig. 91 that this pressure increases as the temperature is raised, although a limit is set by the critical point.

Suppose a quantity of liquid is sealed into a vessel, leaving a free space not filled by the liquid; it is obvious that some of the latter will evaporate until exactly the same condition of equilibrium is attained as would be reached if the vapor alone were taken and partially condensed at the same temperature. Every liquid may, therefore, be regarded as having a tendency to evaporate until the pressure of vapor above it is equal to the equilibrium value, i.e., the vapor pressure of the liquid, for the given temperature. If vaporization occurs into a closed space the vapor will soon attain the equilibrium pressure for the given temperature, but if the space is open, as when water is exposed to the air, evaporation will continue until no more liquid remains.

Kinetic Theory of Vapor Pressure.—The molecules of a liquid, like those of a gas, may be supposed to be in continual motion, and since a liquid has an open surface there is a possibility that some of the molecules will penetrate the surface and enter the space above the liquid. If this space is limited the molecules will steadily accumulate in it, and there will be a tendency, which becomes greater as their number increases, for them to return to the liquid phase. A condition of equilibrium will evidently be attained at each temperature, when the number of molecules leaving the liquid to enter the space above it is equal to the number re-

turning in a given time. The space over the liquid is now saturated with vapor, and the pressure it exerts will be the vapor pressure of the liquid at the given temperature. Consideration shows that this pressure must be independent of the quantity of liquid, or of the space occupied by the vapor. The number of molecules leaving the liquid at a given temperature will be proportional to the area of the open surface, but the number returning will also be proportional to this area and to the pressure of the vapor (see p. 278). The equilibrium pressure will, therefore, depend only on the temperature.

At equilibrium the number of molecules in unit volume of the liquid is greater than in the saturated vapor, the disparity becoming smaller as the critical temperature is approached. The reason for this difference is that before a molecule can leave the surface of a liquid it must overcome the forces of cohesion, resulting from the considerable attraction exerted by the other molecules (cf. p. 299). Only molecules possessing sufficient energy are therefore able to vaporize, and the average potential energy of the molecules in the vapor phase is greater than the mean value in the liquid phase. If n_V and n_L represent the number of molecules or moles per unit volume of vapor and liquid respectively, then it follows (p. 273) that

$$\frac{n_V}{n_L} = e^{-L_i/RT}, \quad (26)$$

where L_i is the mean potential energy difference per mole. As the temperature is raised the ratio n_V/n_L increases, and since n_L does not vary appreciably with temperature, the molecular concentration in the gas phase, and hence the vapor pressure, increases. This matter will be considered quantitatively on p. 452.

Latent Heat of Vaporization.—Since molecules of vapor possess, on the average, more potential energy than those in the liquid in equilibrium with the vapor, it is clear that the process of evaporation must be accompanied by an absorption, and condensation by an evolution, of energy. The heat absorbed when a definite quantity, e.g., 1 g. or 1 mole, of liquid is vaporized is called the **latent heat of vaporization** per g. or per mole, respectively. If the process occurs at constant volume then the heat change is equal to the energy required to overcome the force of attraction in the liquid, and is consequently represented by the quantity L_i in (26); this is called the *internal* latent heat of vaporization per mole. Measurements of latent heat are not generally made at constant volume, but at the vapor pressure of the system. In these circumstances there must be added the energy required to perform the external work resulting from the change of volume,* that is, $p(V_V - V_L)$, where p is the vapor pressure, and V_V and V_L are the orthobaric molar volumes of vapor and liquid, respectively. The total molar latent heat of vaporization L_o is, therefore,

* The external work done is equal to the product of the pressure and the increase of volume (p. 183).

given by

$$L_e = L_i + p(V_v - V_L). \quad (27)$$

An approximate form of this relationship may be obtained by neglecting the volume of the liquid, with respect to that of the vapor, so that the external work term becomes pV_v . If the vapor is assumed to obey the ideal gas laws, this quantity is equal to RT ; hence, for 1 mole,

$$L_e \approx L_i + RT. \quad (28)$$

It may be readily shown that both L_e and L_i must vary with temperature. The orthobaric densities of liquid and saturated vapor become closer as the temperature is raised and are identical at the critical point; it is evident, therefore, that the differences between n_v and n_L , and between V_v and V_L , become smaller as the temperature is raised, so that from (26) and (27) it follows that both L_i and L_e must decrease with increasing temperature. At the critical point it is obvious that both total and internal latent heats must be zero; this is another aspect of the lack of distinction between liquid and vapor in the critical state.

Latent heats of vaporization are measured either by supplying a known amount of heat electrically to the liquid and determining the loss in weight by evaporation, or by finding the rise of temperature resulting from the condensation of a definite quantity of vapor. It is the general practice to tabulate latent heats, generally per gram, determined at the normal boiling point of the liquid, that is, at a pressure of 1 atm.

Vapor Pressure and Boiling Point.—If the pressure above a liquid is adjusted to a definite value, say 1 atm., it is possible to raise the temperature of the liquid until its vapor pressure is equal to the arbitrary external pressure. At this point the liquid can evaporate freely, and bubbles of vapor are seen to form within it; this is, of course, the boiling point of the liquid which can consequently be defined as *the temperature at which the vapor pressure is equal to the external pressure*, generally 1 atm. If the external pressure is decreased, as in distillation under reduced pressure, the boiling point of the liquid will obviously be lowered, whereas increase of external pressure results in a rise of the boiling point.

Vapor Pressure and External Pressure.—The assumption has been made hitherto that the only pressure exerted on the liquid is that of the vapor; if the liquid is subjected to an additional external pressure, however, the vapor pressure will be changed. The external pressure may rise because of the presence of an inert gas in the vapor phase, or as a result of the hydrostatic pressure of a column of the liquid itself (cf. p. 668). The variation of the vapor pressure with the pressure at the liquid surface may be derived thermodynamically.

It will be shown later (p. 477) that when several phases are in equilibrium every component has the same chemical potential in each phase; consequently when a pure liquid, under a total pressure P , is in equilibrium with its vapor at a partial pressure p , the free energy per mole will be the same in both phases. If the pressures are altered, the system still remaining in equilibrium, the change

in free energy must also be the same in both phases. Suppose the external pressure is increased by a small amount dP to $P + dP$, and the equilibrium vapor pressure is consequently changed by dp to $p + dp$, then at constant temperature the corresponding free energy changes [see equation (127), p. 232] for 1 mole of liquid (dF_L) and vapor (dF_V) are given by

$$dF_L = V_L dP \quad \text{and} \quad dF_V = V_V dp,$$

where V_L and V_V are the molar volumes of liquid and vapor, under the pressures P and p , respectively. As shown above, these two free energy changes must be equal; hence, at a definite temperature,

$$\frac{dp}{dP} = \left(\frac{\partial p}{\partial P} \right)_T = \frac{V_L}{V_V}, \quad (29)$$

so that the influence of external pressure on vapor pressure is determined by the relative molar, or specific, volumes of the liquid and vapor. Since the latter is much larger than the former, the effect of external pressure is relatively small. To obtain an approximate indication of the change in vapor pressure, the vapor may be supposed to obey the ideal gas laws, so that V_V can be replaced by RT/p , giving

$$\frac{dp}{p} = \frac{V_L}{RT} dP. \quad (30)$$

Since V_L does not vary appreciably with pressure, liquids being almost incompressible, it may be assumed constant; integration of this equation, between the limits of P_1 and P_2 for the external pressure, the corresponding vapor pressures being p_1 and p_2 , then gives

$$\ln \frac{p_1}{p_2} = \frac{V_L(P_1 - P_2)}{RT}. \quad (31)$$

For an increase of 1 atm. pressure for example, the vapor pressure of water, at ordinary temperatures, can thus be shown to be increased by about 0.1 per cent; the *fractional* increase is less the higher the temperature. As a general rule, therefore, the effect of external pressure is ignored, although the matter has a theoretical interest in connection with osmosis (p. 668).

Measurement of Vapor Pressure.—The procedures most frequently used for studying vapor pressures fall into three categories, viz., (i) static, (ii) dynamic, and (iii) transpiration (gas saturation). The static method in its simplest form was used by J. Dalton (1803) and others; it was improved by H. V. Regnault (1847) and later by W. Ramsay and S. Young (1887). Two barometer tubes are used, one of which is for purposes of comparison; the liquid under investigation, which must be quite free from dissolved air, is introduced into the other, until the space above the mercury is saturated with vapor, as shown by a small quantity of liquid remaining on the surface of the mercury. The difference in levels of the mercury in the two tubes gives the vapor pressure of the liquid at the experimental temperature. The measurements can be made at a series of temperatures by surrounding the barometer tubes with a heating jacket.

In the modern forms of the static method the experimental liquid is placed in a well evacuated bulb which can be connected to a suitable mercury manometer, or to a glass or quartz pressure gauge, of the spiral or sickle type (cf. p.

318), if the vapor attacks mercury; the pressure produced at different temperatures can then be measured directly.

A convenient device for the direct measurement of vapor pressures is the *isotenoscope* of A. Smith and A. W. C. Menzies (1910) shown in Fig. 96; the bulb *A* about 2 cm. diameter is half-filled with the liquid under investigation, which is also placed in the U-tube *B*, 3 to 5 cm. in height. The apparatus is immersed in a constant temperature bath, and after any air present has been boiled out, the pressure in a regulator vessel attached to *C* is adjusted until the level of the liquid is the same in both limbs of *B*; the pressure in the regulator, measured by an attached manometer, is then equal to the vapor pressure of the liquid.

The principle involved in the dynamic method for the measurement of vapor pressure is the following: instead of fixing the temperature and observing the pressure, the external pressure is fixed and the temperature at which the liquid boils is then measured. In accordance with the definition of boiling point, given above, this represents the temperature at which the liquid has a vapor pressure equal to the external pressure. In its simplest form the method has been used for studying various substances including metals. A convenient experimental procedure has been described by W. Ramsay and S. Young (1884) for relatively volatile liquids. A boiling tube is fitted with a rubber stopper carrying a thermometer with its bulb covered by absorbent cotton, a dropping funnel containing the liquid under investigation, and an outlet tube for connection to a large bottle acting as a pressure reservoir. The pressure can be adjusted by means of a pump and a side-tube for the admission of air if required. With the aid of a mercury manometer, the pressure in the apparatus is adjusted to a definite value at which it must be kept constant. The tube and its contents are then warmed until the conditions are such that the thermometer reading does not change when a small quantity of liquid is allowed to flow out of the funnel and moisten the cotton. The liquid surrounding the thermometer is then in equilibrium with vapor at the pressure maintained in the apparatus, and the temperature recorded is the boiling point of the liquid at this pressure. By admitting or removing air, the pressure in the apparatus can be altered and the new boiling point determined. In this manner the vapor pressures for a series of temperatures can be measured.

In the method of A. Smith and A. W. C. Menzies (1910) a narrow tube with a bulb, containing the experimental liquid, at one end is attached to a thermometer and immersed in a boiling tube containing a liquid of high boiling point, e.g., paraffin (Fig. 97); the boiling tube is connected to a pressure reservoir and manometer as described above. The temperature of the paraffin is raised to

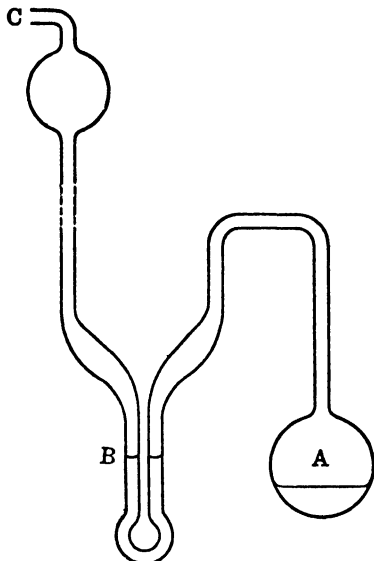


FIG. 96. Isotenoscope (Smith and Menzies)

a convenient value and the pressure in the apparatus reduced until the liquid in the bulb commences to boil and a stream of bubbles escapes. Air is now cautiously admitted to the apparatus until the bubbles are just about to cease; the temperature and pressure are then recorded, the former being the boiling point for that particular pressure, after applying a correction for the hydrostatic pressure of the column of paraffin above the bulb-tube. The observations are repeated at a number of other temperatures.

The transpiration method in a simple form was employed by H. V. Regnault (1845) to measure the vapor pressure of water; it has since been used for various substances, and especially for solutions (p. 630) and salt hydrates (p. 781). If dry air, or an inert gas, is bubbled slowly through a liquid, in a suitable saturator, at constant temperature, the gas will become saturated with the vapor of the liquid, and the partial pressure of the latter in the resulting mixture will be equal to the vapor pressure. Assuming that Dalton's law of partial pressures holds and that the vapor behaves ideally, then

$$pV = \frac{w}{M} RT, \quad (32)$$

where p is the partial pressure of the vapor, V is the total volume of gas and vapor formed in the saturator, w is the weight of liquid vaporized, determined from the loss in weight of the saturator or the gain in weight of an absorber, and M is the molecular weight of the vapor. An important source of error in the application of the transpiration method lies in the measurement of the large volume of air; this has been overcome in the modification of J. N. Pearce and R. D. Snow (1927). The gas used is a mixture of oxygen and hydrogen produced electrolytically, and its amount is known very precisely from measurements of the silver deposited in a coulometer in the circuit (see p. 886). If p is the pressure of the vapor and P the total pressure of the gas and vapor, equal to the corrected atmospheric pressure, then if Dalton's law is applicable

$$p = Pn_1/(n_1 + n_2),$$

where n_1 is the number of moles of vapor and n_2 the total number of moles of hydrogen and oxygen in the saturated vapor; n_1 is known from the mass of liquid vaporized, and n_2 from the quantity of electricity passed, and so p can be evaluated.

The vapor pressures of a number of metals have been measured by the transpiration method, the inert gas, e.g., hydrogen, being led over the surface of the molten metal contained in a weighed boat. Under these conditions it is not certain that the gas is completely saturated; determinations of the loss of weight are therefore made with the gas flowing at different rates, and the results extrapolated to zero rate of flow. If n_G moles of gas are passed and n_M moles of metal are vaporized, the vapor pressure may be taken as equal to $Pn_M/(n_G + n_M)$, where P is the gas pressure. The vapor pressures obtained by the transpiration method are for an external pressure equal to that of the

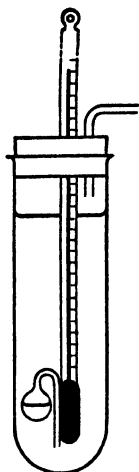


FIG. 97. Vapor pressure measurement (Smith and Menzies)

gas used, generally 1 atm., but the influence of this additional pressure may be neglected (p. 446).

A method for the determination of low vapor pressures, e.g., of metals, is based on the measurement of the rate of evaporation from the surface of the liquid (I. Langmuir, 1913; P. Harteck, 1928; A. L. Marshall, 1937). It has been seen (p. 278) that the mass of gas μ striking 1 sq. cm. of a surface per sec. is given by

$$\mu = p(M/2\pi RT)^{1/2}, \quad (33)$$

where p is the pressure and M the molecular weight of the gas. When a liquid and saturated vapor are in equilibrium the same equation must represent both the rate at which molecules return to the liquid and that at which they evaporate (p. 278), p being the vapor pressure of the liquid. Since the saturated vapor has the highest pressure possible for the liquid-vapor system at the given temperature, the rate of return of molecules to the liquid is the maximum; consequently, if p is the vapor pressure, (33) represents the *maximum* rate of vaporization from the liquid surface. By measuring the loss of weight, in a given time, of liquid of known surface area when heated in a high vacuum, the maximum rate of evaporation can be determined, and hence the vapor pressure evaluated.

In another procedure (M. Knudsen, 1909; A. C. G. Egerton, 1917, 1923; W. H. Rodebush, 1925), sometimes called the "effusion method," the rate is determined at which the vapor passes through a hole whose diameter is small in comparison with the mean free path of the molecules under the experimental conditions (cf. p. 278). This rate is equivalent to that at which the molecules, at the vapor pressure of the liquid, would strike a surface having the same area as the hole; hence by means of (33) the vapor pressure can be calculated.*

The vapor pressures of a number of liquids at various temperatures have been collected in Table 68.

TABLE 68. VAPOR PRESSURES OF LIQUIDS (MM. OF MERCURY)

Temp.	Water	Ethyl alcohol	Acetic acid	Acetone	Ethyl acetate	Carbon tetrachloride	Chloroform	Aniline	Chlorobenzene
10° c.	9.21	23.6	—	115.6	42.8	56.0	100.5	—	4.86
20°	17.53	43.9	11.7	184.8	72.8	91.0	159.6	—	8.76
30°	31.82	78.8	20.6	282.7	118.7	143.0	246.0	—	15.45
40°	55.32	135.3	34.8	421.5	186.3	215.8	366.4	—	26.00
50°	92.51	222.2	56.6	612.6	282.3	317.1	526.0	2.4	41.98
60°	149.4	352.7	88.9	(860.6)	415.3	450.8	739.6	5.7	65.54
70°	233.7	542.5	136.0	—	596.3	622.3	(1019)	10.6	97.90
80°	355.1	(812.6)	202.3	—	(832.8)	(843)	—	18.0	144.75
90°	525.8	—	293.7	—	—	—	—	29.2	208.35
100°	760.0	—	417.1	—	—	—	—	45.7	292.75
Boiling Point	100°	78.32°	118.5°	56.10°	77.15°	76.75°	60.9°	184.1°	132.0°

Vapor Pressure and Temperature.—The variation of the vapor pressure of a liquid with temperature is invariably represented by a curve of

the type shown in Fig. 98; the highest temperature at which vapor pressure can be measured is, of course, the critical point. In view of the significance of the boiling point, it is apparent that the same curve represents the influence of pressure on the boiling point of a liquid. The curve is evidently logarithmic in form, as was realized by J. Dalton (1804), and many attempts have been made to devise empirical equations to represent the influence of temperature on vapor pressure; the subject is best considered, however, with the aid of an important thermodynamic equation which will now be deduced.

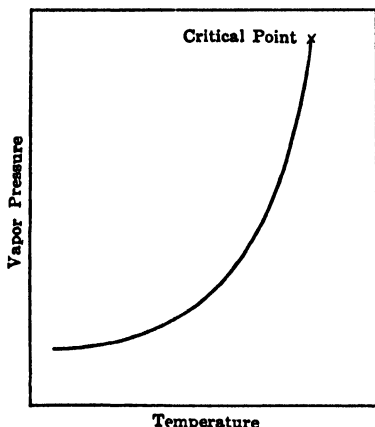


FIG. 98. Vapor pressure-temperature curve

the influence of temperature on vapor pressure; the subject is best considered, however, with the aid of an important thermodynamic equation which will now be deduced.

The Clapeyron-Clausius Equation.

—It has been shown on p. 236 [equation (144)] that for a closed system

$$\left(\frac{\partial P}{\partial T}\right)_v = \left(\frac{\partial S}{\partial V}\right)_T. \quad (34)$$

This relationship can be applied to any closed univariant system (cf. p. 475) of two or more phases, but its immediate interest is in connection with a system of two phases of the same substance. Consider, for example, a system consisting of liquid and vapor in equilibrium.

The pressure p , which is the vapor pressure, is dependent on the temperature, but is independent of the volume of the liquid and vapor; hence dp/dT may be written in place of $(\partial P/\partial T)_v$. Suppose 1 mole of liquid is vaporized isothermally and reversibly at a definite temperature T , so that liquid and vapor remain at equilibrium during the process; the increase of heat content ΔH is equal to the molar heat of evaporation, and since this is taken up reversibly, it follows that (p. 222)

$$\Delta S = \Delta H/T. \quad (35)$$

Further, if ΔV is the increase of volume accompanying the vaporization of 1 mole of liquid, then $\Delta S/\Delta V$ has a constant value for each temperature, and so it may be substituted for $(\partial S/\partial V)_T$. Making the simplifications in (34) and utilizing (35), it follows that

$$\frac{dp}{dT} = \frac{\Delta H}{T\Delta V}, \quad (36)$$

which is a form of the equation first deduced by B. P. E. Clapeyron (1834), and later extended by R. Clausius (1850); it is therefore called the **Clapeyron-Clausius equation**. It can be derived in a number of different ways, although they are all dependent, like the one given, on the second law of thermodynamics.

Although the supposition was made above that the system consists of liquid and vapor, the same equation can be derived in an exactly analogous manner for any equilibrium between two phases, e.g., solid and liquid, solid and vapor, or two crystalline forms, of the same substance. In every case ΔH is the heat absorbed in the transition and ΔV is the accompanying increase of volume.

For the particular case of liquid-vapor equilibrium, ΔH is equal to L_e , the molar heat of evaporation, and the volume change ΔV is $V_v - V_L$, where V_v and V_L are the molar volumes of vapor and liquid, respectively; hence, for the variation of the vapor pressure with temperature, (36) may be written

$$\frac{dp}{dT} = \frac{L_e}{T(V_v - V_L)} \quad (37)$$

$$= \frac{l_e}{T(v_v - v_L)}, \quad (38)$$

where l_e is the latent heat per gram, and v_v and v_L are the *specific* volumes of vapor and liquid.

The Clapeyron-Clausius equation is of value in a number of connections; for example, if dp/dT is known from experiment at any temperature, it is possible to calculate the heat of vaporization at that temperature. This application will be illustrated by means of the well-established data for water. Between 99.5° and 100.5° c., dp/dT is 2.717 cm. of mercury per degree,

$$\therefore \frac{dp}{dT} = 2.717 \times 13.59 \times 981 \text{ dynes per sq. cm. per deg.}$$

The specific volumes of water vapor and of liquid water at 100° are 1674 cc. and 1 cc. respectively, and so $v_v - v_L = 1673$ cc. Substitution in (37) at 100° c., i.e., $T = 373.2^\circ$, would give l in ergs, but as latent heats are generally expressed in calories, it is necessary to divide by 4.184×10^7 (see p. 184); hence

$$l_e = \frac{373.2 \times 1673 \times 2.717 \times 13.59 \times 981}{4.184 \times 10^7} = 540 \text{ cal. per g.}$$

The experimental value for the latent heat of vaporization of water at 100° c. is 539 cal. per g. It is thus possible to calculate the latent heats of vaporization at a series of temperatures, provided the vapor pressure-temperature curve has been established experimentally. The results become less accurate as the critical point is approached because $v_v - v_L$ becomes almost zero.

If, on the other hand, the latent heat is known at any temperature, it is possible to evaluate dp/dT , the rate of increase of vapor pressure with temperature, at that temperature. From the fact that vapor pressure-temperature and boiling point-pressure curves are identical, it will be seen that dT/dp represents the influence of pressure on the boiling point. Taking the latent heat of vaporization of water, for example, as 539 cal. at 100° c., dT/dp is calculated as 0.37 degrees per cm. of mercury; hence water will boil at about 100.37° c. at a pressure of 77 cm., and 99.63° c. at 75 cm. pressure.

Vapor Pressure Equations.—If the temperature is not near the critical value, the volume of the liquid will be small and can be neglected in

comparison with that of the vapor; the Clapeyron-Clausius equation then takes the form

$$\frac{dp}{dT} = \frac{L_v}{TV_v}. \quad (39)$$

Further, in regions well below the critical temperature, the vapor pressure is relatively small and the ideal gas laws may be assumed to be applicable, so that

$$pV_v = RT,$$

since V_v is the molar volume of the vapor; hence RT/p can be substituted for V_v in (39) giving

$$\frac{dp}{dT} = \frac{L_v p}{RT^2}. \quad (40)$$

Using the mathematical identity

$$\frac{d \ln p}{dT} = \frac{1}{p} \cdot \frac{dp}{dT},$$

it follows that

$$\frac{d \ln p}{dT} = \frac{L_v}{RT^2}. \quad (41)$$

Since L_v is the heat absorbed at constant pressure it may be put equal to ΔH_v , the increase of heat content (p. 206); thus,

$$\frac{d \ln p}{dT} = \frac{\Delta H_v}{RT^2}. \quad (42)$$

This approximate result, derived by Clausius, is identical in form with the van't Hoff equation (p. 828), which is applicable to equilibria of all kinds; (41) and (42) may, therefore, be regarded as one aspect of a more general relationship.

It is of interest to observe that (41) can also be derived from the kinetic considerations already discussed; (26) on p. 444 may be written

$$\ln \frac{n_v}{n_L} = - \frac{L_i}{RT}, \quad (43)$$

where n_v and n_L are the number of moles per unit volume of vapor and liquid, respectively. Since the coefficient of expansion of a liquid is small, the number of moles per unit volume (n_L) may be taken as independent of temperature; differentiation of (43) with respect to temperature then gives

$$\frac{d \ln n_v}{dT} = \frac{L_i}{RT^2}. \quad (44)$$

If the unit volume is 1 liter, then $1/n_v$ is the molar volume in liters, i.e., V_v , and if the ideal gas laws are obeyed this is equal to RT/p , where p is the vapor pressure; hence,

$$\ln n_v = \ln p - \ln RT,$$

and (44) becomes

$$\frac{d \ln p}{dT} = \frac{L_i + RT}{RT^2} = \frac{L_o}{RT^2},$$

since $L_i + RT = L_o$, the total latent heat for an ideal gas, provided the molar volume of the liquid is small in comparison with that of the vapor (p. 445).

The advantage of (41) is that it can be readily integrated if it is assumed that the latent heat L_o is constant over a range of temperature; this can only be true if the range is small. If the temperature limits are T_1 and T_2 , and the corresponding vapor pressures are p_1 and p_2 , respectively, integration gives

$$\ln \frac{p_2}{p_1} = -\frac{L_o}{R} \left(\frac{1}{T_2} - \frac{1}{T_1} \right) = \frac{L_o}{R} \left(\frac{T_2 - T_1}{T_1 T_2} \right) \quad (45)$$

or

$$\log \frac{p_2}{p_1} = \frac{L_o}{4.576} \left(\frac{T_2 - T_1}{T_1 T_2} \right) \quad (46)$$

with R equal to 1.987 cal. This equation, like the original Clapeyron-Clausius form, can be employed to calculate latent heats from vapor pressure data, and *vice versa*, and also to determine the influence of pressure on the boiling point; the results obtained with (46) are, as might be expected, less accurate.

The general integral of (41), assuming L_o to be constant, is

$$\ln p = -\frac{L_o}{RT} + c; \quad (47)$$

$$\therefore \log p = c - \frac{a}{T}, \quad (48)$$

where a and c are constants for the given substance. According to (48) the plot of $\log p$ against $1/T$ should be a straight line, but this is only true over a limited range of temperature because of the variation in the latent heat. This variation may be expressed (approximately, see p. 462) by the Kirchhoff equation in the form

$$\frac{dL_o}{dT} = C_V - C_L, \quad (49)$$

where C_V and C_L are the molar heat capacities of vapor and liquid, respectively, at constant pressure. At any temperature, therefore,

$$L_o = L_0 + \int_0^T (C_V - C_L) dT = L_0 + \int_0^T \Delta C_P dT, \quad (50)$$

or

$$L_o(T_2) = L_o(T_1) + \int_{T_1}^{T_2} \Delta C_P dT, \quad (50a)$$

where ΔC_P is equal to the difference in the molar heat capacities of vapor and

liquid. The corresponding form of (41) derived from (50) is then

$$\frac{d \ln p}{dT} = \frac{L_0 + \int_0^T \Delta C_P dT}{RT^2} \quad (51)$$

and on integration (cf. p. 860)

$$\ln p = -\frac{L_0}{RT} + \int \frac{dT}{RT^2} \int_0^T \Delta C_P dT + \text{const.} \quad (52)$$

$$= -\frac{L_0}{RT} + \int_0^T \frac{\Delta C_P}{RT} dT + \text{const.} \quad (53)$$

This equation gives the variation of vapor pressure with temperature, but its application requires a knowledge of the influence of temperature on the heat capacities of vapor and liquid. If the simplifying assumption is made that ΔC_P is independent of temperature, so that it may be replaced by the constant α ,* then (50) becomes

$$L_e = L_0 + \Delta C_P T = L_0 + \alpha T \quad (54)$$

and (53) reduces to

$$\ln p = -\frac{L_0}{RT} + \frac{\Delta C_P}{R} \ln T + \text{const.} \quad (55)$$

or

$$\log p = -\frac{a}{T} + b \log T + c, \quad (56)$$

where a , b and c are constants. Equations of the type of (56) were proposed empirically by G. Kirchhoff (1858) and others, in various forms, and have been found to be applicable to many substances over a large range of temperature.

The empirical relationship

$$\log \frac{p}{P_c} = k \left(1 - \frac{T_c}{T} \right), \quad (57)$$

where k is a constant, and P_c and T_c are the critical pressure and temperature of the given substance, was suggested by J. D. van der Waals (1881). Since P_c and T_c are constants this equation may be readily reduced to the form of (48), and must consequently be subject to the same limitations. The special interest of (57), however, lies in the fact, deduced from the law of corresponding states, that k should be a universal constant; this has been found approximately true, the value being close to 3.0 for a number of liquids.

Comparison of Different Substances: The Ramsay-Young and Dühring Rules.—According to the approximate equation (48) the vapor pressures of two substances A and B can be represented by

$$\log p_A = -\frac{a_A}{T} + c_A \quad \text{and} \quad \log p_B = -\frac{a_B}{T} + c_B.$$

* Since L_e is zero at the critical point it must decrease with increasing temperature, at least as the critical temperature is approached; in these circumstances ΔC_P and α must be negative and the heat capacity of the liquid should be, as it almost invariably is, greater than that of the vapor at the same temperature.

If these two liquids have the same vapor pressure p at temperatures T_A and T_B , respectively, then

$$-\frac{a_A}{T_A} + c_A = -\frac{a_B}{T_B} + c_B;$$

$$\therefore \frac{T_A}{T_B} = \frac{T_A}{a_B} (c_B - c_A) + \frac{a_A}{a_B}. \quad (58)$$

At the temperatures T'_A and T'_B respectively, the substances now have the same vapor pressure p' ; hence,

$$\frac{T'_A}{T'_B} = \frac{T'_A}{a_B} (c_B - c_A) + \frac{a_A}{a_B}, \quad (59)$$

and subtraction of (59) from (58) gives

$$\frac{T_A}{T_B} - \frac{T'_A}{T'_B} = c(T_A - T'_A), \quad (60)$$

where c is a constant. This equation was derived empirically by W. Ramsay and S. Young (1885) who showed that c was small, and that if A and B are related chemically, e.g., two esters or chloro- and bromobenzenes, c is negligible, and (60) becomes

$$T_A/T_B = T'_A/T'_B. \quad (61)$$

Ramsay and Young found this equation to hold closely for a number of related substances so that for these compounds T_A/T_B is a constant at all vapor pressures; the ratio of the boiling points of two similar substances is, therefore, the same no matter what the external pressure. The Ramsay and Young equation is consequently of value for the comparison of vapor pressures at different temperatures, and of boiling points at different pressures, for compounds of the same type.

By slight rearrangement (61) gives

$$\frac{T'_A - T_A}{T'_B - T_B} = \frac{T_A}{T_B} = \frac{T'_A}{T'_B}; \quad (62)$$

$$\therefore \frac{T'_A - T_A}{T'_B - T_B} = \text{constant}, \quad (63)$$

and this constant should be independent of the pressure; this is the rule discovered by U. Dühring (1878) for the boiling points of two substances under different pressures.¹⁰

For associated substances containing a hydroxyl group, e.g., alcohols and carboxylic acids, the relationships given above do not apply; the simple equation (48) is not satisfactory, mainly because of the expected large variation of latent heat with temperature, and the value of c in (60) is not constant.

Boiling Point and Critical Temperature.—If the pressure is put equal to 1 atm., then T is the boiling point T_b , and the van der Waals vapor pressure equation (57) becomes

$$\log \frac{1}{P_o} = k \left(1 - \frac{T_o}{T_b} \right); \quad (64)$$

$$\therefore \frac{T_b}{T_o} = \frac{k}{k + \log P_o}. \quad (65)$$

The critical pressures of most substances are about 50 to 100 atm., so that $\log P_o$ is approximately 2; since k has been found to be 3 for many substances, it is seen that

$$\frac{T_b}{T_o} \approx 0.6. \quad (66)$$

The relationship between the normal boiling point of a liquid and its critical temperature, implied by this equation, was pointed out by C. M. Guldberg (1890) and P. A. Guye (1890). The data in Table 69 show that

TABLE 69. RATIO OF BOILING POINT TO CRITICAL TEMPERATURE

Substance	T_b	T_o	Ratio	Substance	T_b	T_o	Ratio
Helium	4.2° K.	5.2° K.	0.81	Ethane	185° K.	305° K.	0.61
Hydrogen	20.4°	33.2°	0.61	<i>n</i> -Pentane	309°	470°	0.66
Nitrogen	77.3°	126.0°	0.61	Benzene	353°	562°	0.63
Oxygen	90.1°	154.3°	0.58	Carbon tetrachloride	350°	556°	0.63
Argon	87.4°	150.7°	0.58	Propyl acetate	375°	549°	0.68
Carbon dioxide	194.6°	304.2°	0.64	Ethyl alcohol	351°	516°	0.68
Ammonia	239.7°	405.5°	0.59	Acetic acid	391°	594°	0.66
Water	373.2°	647°	0.58	Iodobenzene	461°	721°	0.62

it is approximately true for a variety of substances, normal and associated, although helium is an outstanding exception. Since the boiling point is approximately a constant fraction of the critical temperature, the former may be regarded as a corresponding temperature (p. 440).¹¹

Heat of Vaporization and Boiling Point.—The simplified form (40) of the Clapeyron-Clausius equation can be written

$$\frac{T}{p} \cdot \frac{dp}{dT} = \frac{L_o}{RT}, \quad (67)$$

and if T and p on the left-hand side are replaced by θT_o and πP_o , respectively (cf. p. 439), it follows that

$$\frac{\theta}{\pi} \cdot \frac{d\pi}{d\theta} = \frac{L_o}{RT}. \quad (68)$$

Further, the van der Waals vapor pressure equation (57) may be put in the form

$$\ln \pi = 2.3k \left(1 - \frac{1}{\theta} \right), \quad (69)$$

and differentiation gives

$$\frac{d \ln \pi}{d\theta} = 2.3 \frac{k}{\theta^2} \quad \text{or} \quad \frac{1}{\pi} \cdot \frac{d\pi}{d\theta} = 2.3 \frac{k}{\theta^2}, \quad (70)$$

and hence

$$\frac{\theta}{\pi} \cdot \frac{d\pi}{d\theta} = 2.3 \frac{k}{\theta}. \quad (71)$$

Comparison of (68) and (71) shows that

$$\frac{L_e}{RT} = 2.3 \frac{k}{\theta} \quad \therefore \frac{L_e}{T} = 6.9 \frac{R}{\theta}, \quad (72)$$

utilizing the fact recorded above that k is 3.0 for many substances. It has been seen that if the temperature is the boiling point T_b , then θ , which is T_b/T_e , is approximately 0.6, and therefore

$$\frac{L_e}{T_b} \approx 23 \quad (73)$$

taking R as 2 cal. deg.⁻¹ mole⁻¹.

Trouton's Law.—The molar latent heat of vaporization of a liquid divided by its boiling point on the absolute scale, at atmospheric pressure, should thus be constant, equal approximately to 23 if the latent heat is expressed in calories.* It should be noted that this quantity is incidentally the change of entropy (p. 224) accompanying the evaporation of 1 mole, and hence the entropy of vaporization should be approximately the same for all substances at atmospheric pressure. If the Ramsay and Young equation (61) is obeyed then these rules should hold at any pressure, although the constant will have a different value. The generalization represented by (73) was first put forward by A. Pictet (1876), and rediscovered by W. Ramsay (1877) and F. Trouton (1884); it is commonly known as **Trouton's law**. It is of interest to record that B. Despretz (1823) found the quantity $L_e/(v_v - v_L)$ to be a constant at the boiling point; if v_L is neglected in comparison with v_v , and the vapor is assumed to obey the ideal gas laws, it can be readily shown that L_e/T_b should be constant. The results in Table 70 indicate how far various substances obey Trouton's law.

TABLE 70. TEST OF TROUTON'S LAW

Substance	L_e cal.	T_b	L_e/T_b	Substance	L_e cal.	T_b	L_e/T_b
Helium	22	4.2°K.	5.2	Ethyl ether	6,466	307°K.	21.1
Hydrogen	216	20.4°	10.6	Benzene	7,497	353°	21.2
Carbon monoxide	1,414	81.6°	17.3	Propyl acetate	8,310	375°	22.2
Nitrogen	1,362	77.3°	17.6	Mercury	14,200	630°	22.6
Oxygen	1,630	90.1°	18.0	Zinc	27,730	1180°	23.5
Methane	1,951	108°	18.0	Potassium			
Sodium	23,300	1155°	20.2	chloride	40,500	1690°	24.0
Carbon				Water	9,700	373°	26.0
disulfide	6,490	319°	20.4	Ethyl alcohol	9,448	351°	26.9

* The same result can be derived theoretically in other ways.

The Trouton rule is evidently only approximate; for nonassociated substances of molecular weight of about 100, and of not too high boiling point, L_v/T_b is approximately 21. The generalization breaks down, however, for associated compounds, e.g., water and alcohol, for substances of high boiling point, which give high values for L_v/T_b , and for those of low molecular weight and low boiling point, for which L_v/T_b is exceptionally low.

Various attempts have been made to improve the Trouton equation by including a term involving the boiling point, e.g., $L_v/T_b = \text{constant} + f(T_b)$, or in other ways. One of the most interesting modifications is that suggested by J. H. Hildebrand (1915) who found that the entropy of vaporization L_v/T was constant for a whole range of substances, varying from nitrogen to zinc, provided it was *measured at the same molecular concentration of the vapor* in every case, instead of at the same pressure of vapor as in the Trouton rule. Hildebrand's method of evaluating L_v/T is to plot $\log p$ against $\log T$ and to determine the slopes of the tangents to the curves when $\log p = \log T - x$, where x is a convenient figure, e.g., 0.5. If the gas laws are obeyed $\log p = \log T + \log Rc$, and so a constant value of x implies a constant concentration of vapor. According to (67) $d \log p / d \log T = L_v/RT$, and so the slope measured gives L_v/RT in every case under such conditions that the concentration of the vapor is constant. Helium shows some deviation from Hildebrand's rule, but that for hydrogen is much less than in Table 70; the behavior of associated liquids, such as water, ammonia and methyl alcohol, is still abnormal, the entropy of vaporization being exceptionally high. Whenever there is any regularity of structure or restriction to movement in the liquid, e.g., due to hydrogen bond formation or to the shape of the molecules, the increase of randomness upon evaporation, and hence the entropy change (cf. p. 229), will be higher than normal; this is evidently the case for associated substances and also for those possessing long-chain molecules.¹²

Boiling Point and Pressure: Craft's Rule.—It has been seen that provided sufficient data are available it is possible to determine the influence of pressure on the boiling point by means of the Clapeyron equation; a simpler method is, however, possible in many instances. If ΔT is the rise of boiling point resulting from an increase of Δp in the pressure, in the vicinity of 1 atm., then

$$\Delta T = \left(\frac{dT}{dp} \right) \Delta p \quad \text{or} \quad \Delta T = \left(\frac{1}{T} \cdot \frac{dT}{dp} \right) T \Delta p, \quad (74)$$

where dT/dp is the value near the normal boiling point. It was shown by W. Ramsay and S. Young (1886) that for related substances the quantity $T(dp/dT)$ was constant, and even for compounds of different types, provided they were not associated, it did not vary by more than about 10 per cent; it follows, therefore, that

$$\Delta T \approx c T \Delta p, \quad (75)$$

where c is approximately constant for most substances. This equation is the basis of the rule introduced by J. M. Crafts (1887) for the correction of boiling points to atmospheric pressure; if t° c. is the observed boiling point at a pressure of p mm., then the correction ΔT to be added to give the boiling point at 760 mm. is

$$\Delta T = c(273 + t)(760 - p), \quad (76)$$

where c is 0.00012. For alcohols, water and carboxylic acids c is 0.00010, and for substances of very low boiling point, e.g., oxygen, nitrogen and ammonia, it is about 0.00014.

Consideration will show that the constancy of $T(dp/dT)$ for various substances is a direct consequence of the Ramsay and Young boiling point law; Craft's rule can, however, be derived in another manner. Combination of Trouton's equation, $L_v/T_b = 21$, and that of Clapeyron in the form of (40), at the normal boiling point, gives

$$\frac{dT}{dp} = \frac{RT}{21p} \quad (77)$$

and so in the vicinity of the boiling point

$$\Delta T = \frac{R}{21p} T \Delta p. \quad (78)$$

Since R is 2 cal., and p is 760 mm., it follows that

$$\Delta T = 0.00012 T \Delta p \quad (79)$$

in agreement with experiment; Craft's rule should, therefore, be applicable to those substances for which the Trouton equation holds.¹³

SOLID-VAPOR EQUILIBRIUM

Vapor Pressures of Solids.—A solid, like a liquid, has a definite vapor pressure at each temperature*; this pressure may be extremely small, but it is nevertheless definite. If the vapor pressure of the solid is high enough it may be measured by the direct method, but as a general rule this is not sufficiently sensitive; the procedures involving either transpiration, determination of loss of weight in a vacuum, or effusion (pp. 448, 449) have been more commonly employed. The principles are identical with those described in connection with the vapor pressures of liquids. The vapor pressure of a solid increases with temperature and the variation can be represented by a curve similar to that for a liquid; it is generally called a **sublimation curve**, the term **sublimation** being used to indicate the direct conversion of solid to vapor, without the intervention of liquid. The change from solid to vapor is accompanied, like the analogous change from liquid to vapor, by an absorption of heat; this is the latent heat of sublimation, represented by L_s cal. per mole, or l_s cal. per g. The internal heat of sublimation may be obtained from this by the method described for liquids (p. 444).

* Some solids cannot exist in contact with vapor; examples are helium below the critical temperature (p. 512) and various forms of ice (p. 473).

As already noted the Clapeyron-Clausius equation can be applied to the solid-vapor equilibrium, the appropriate form being

$$\frac{dp}{dT} = \frac{L_s}{T(V_v - V_s)}. \quad (80)$$

It can be readily deduced by the method given on p. 450 by considering a system of solid and its vapor in equilibrium. The chief use of this equation is that it permits the latent heat of sublimation to be calculated from the slope of the sublimation curve at any particular temperature. On the other hand, if the latent heat is known (p. 461) the influence of temperature on the sublimation pressure may be evaluated. By neglecting the volume of the solid with respect to that of the vapor, and assuming the latter behaves as an ideal gas, the following equation, which is analogous to (41), may be derived

$$\frac{d \ln p}{dT} = \frac{L_s}{RT^2} = \frac{\Delta H_s}{RT^2}, \quad (81)$$

where ΔH_s is equal to L_s , and represents the heat absorbed, at constant pressure, per mole of solid vaporized. The equation may be integrated with results exactly parallel to those already given for liquids, e.g., equations (46), (48) and (56). The subject of the vapor pressure of solids will be considered again later in various connections (pp. 469, 860).

SOLID-LIQUID EQUILIBRIUM

Fusion and Solidification.—When a pure crystalline solid is heated a temperature is reached at which it changes sharply into liquid; this is known as the **melting point**, and it has a definite value depending on the external pressure. If the liquid is cooled solidification will occur at the same temperature for the given pressure; for a pure substance, therefore, the melting and freezing points are identical.* In other words, for every pressure there is a definite temperature at which a pure solid and liquid can be in equilibrium together; this generalization was deduced theoretically by J. Thomson (1849) and verified experimentally in 1850 by W. Thomson (Lord Kelvin) and by R. Bunsen. It is thus possible to plot a curve representing the conditions of temperature and pressure under which the solid and liquid are in equilibrium; it is frequently called the **melting-point curve**, as it represents the influence of pressure on the melting point. The suggestion has been made that the melting point T_m , like the boiling point, is a corresponding temperature, and that T_m/T_c has an approximately constant value of 0.44; the freezing point of any

* A liquid can sometimes be supercooled, that is the temperature reduced below the freezing point without solid appearing. The reverse process, i.e., suspended transformation in the change from solid to liquid, is rare; it has been observed with certain polymers, when the melting process probably involves the breaking of bonds.

substance should thus be about 0.7 times the boiling point on the absolute scale. These relationships are only very approximate. As with evaporation and sublimation the process of fusion, or melting, is accompanied by an absorption of heat; this is the latent heat of fusion, represented by the symbols L_f and l_f per mole or per g., respectively. Since the volume change accompanying fusion is relatively small the internal latent heat differs very little from the total value measured at constant pressure.

The Clapeyron-Clausius equation applied to the process of fusion is written in the inverted form; thus,

$$\frac{dT}{dP} = \frac{T(V_L - V_S)}{L_f} = \frac{T(v_L - v_S)}{l_f} \quad (82)$$

when it represents the influence of external pressure P on the melting point of the solid. The deduction is readily made by the method of p. 450, a system containing solid and liquid in equilibrium being considered.

The use of (82) may be illustrated by the data for the ice-water system; at 0°C. , i.e., $T = 273.2^\circ$, $v_L = 1.0001 \text{ cc.}$, $v_S = 1.0907 \text{ cc.}$, and $l_f = 79.8 \text{ cal. per g.}$, i.e., $79.8 \times 4.184 \times 10^7 \text{ ergs per g.}$ Insertion of these values in the equation gives the change in melting point for 1 dyne per sq. cm. of the pressure. Multiplication by $1.013 \times 10^6 \text{ dynes per sq. cm.}$ then gives the increase in melting point per atm. increase of pressure; that is,

$$\frac{dT}{dP} = \frac{273.2 \times (-0.0906) \times 1.013 \times 10^6}{79.8 \times 4.184 \times 10^7} = -0.0075^\circ \text{ per atm.}$$

An increase of 1 atm. in the pressure should thus *lower* the melting point of ice by 0.0075° ; this is in good agreement with the result of experiment. The explanation of the negative sign of dT/dP is that v_S is greater than v_L ; that is, the specific volume of the solid is greater than that of the liquid or, in other words, the solid has a lower density than the liquid at the melting point. Water is one of the few substances exhibiting this unusual type of behavior. For most substances the solid has a higher density than the liquid with which it is in equilibrium, and consequently increase of pressure brings about an increase in the melting point. The conclusions drawn from the Clapeyron-Clausius equation are in agreement with those arrived at by the application of the Le Chatelier principle (p. 831). Some calculated and observed values of the change in melting point for 1000 atm. increase of pressure are collected in Table 71; the

TABLE 71. INFLUENCE OF PRESSURE ON MELTING POINT

Substance	Melting Point	l_f cal./g.	$v_L - v_S$	ΔT for 1000 atm.	
				Calc.	Obs.
Water	273.2°K.	79.8	-0.0906	-7.5	-7.4
Acetic acid	289.8°	44.7	+0.01595	25.0	24.4
Tin	505°	14.0	+0.00389	3.40	3.28
Bismuth	544°	12.6	-0.00342	-3.56	-3.55

agreement is seen to be good. The results show, incidentally, that as a general rule very large pressures are required to produce an appreciable change of melting point.

Latent Heat of Fusion.—The latent heat of fusion is related to the latent heats of vaporization and sublimation; the first law of thermodynamics leads to the equation

$$L_s = L_l + L_v, \quad (83)$$

since the same amount of heat must be absorbed in the conversion of a given quantity of solid directly to the vapor (L_s) as would be required for the change in two stages, first from solid to liquid (L_l) and then from liquid to vapor (L_v), at the same temperature and pressure.

The influence of temperature on the latent heat of fusion is often expressed by a form of the Kirchhoff equation; thus,

$$\frac{dL_l}{dT} = C_L - C_s, \quad (84)$$

where C_L and C_s are the molar heat capacities of liquid and solid, respectively. It must be emphasized, however, that the Kirchhoff relationship is only applicable to processes taking place at constant pressure or constant volume, whereas changes of state are not of this type. If the temperature at which a solid is melted or a liquid vaporized is changed, there must be a corresponding change of pressure, and so (84) cannot hold. By a more exact thermodynamic treatment it can be shown that for the liquid-vapor change (49) is at least approximately true, but for the fusion process (84) should be replaced by

$$\frac{dL_l}{dT} = C_L - C_s + \frac{L_l}{T}. \quad (85)^*$$

Since the melting point changes with the applied pressure, so also will the latent heat; the direction of the change obviously depends on the sign of dT/dP , for this determines whether the melting temperature is raised or lowered by increase of pressure.

Entropy and Mechanism of Fusion.—According to P. Walden (1908) a rule analogous to that of Trouton for vaporization applies to fusion; he claimed that for many compounds the quantity L_l/T_m , i.e., the entropy of fusion, is about 13 E.U. (cal. deg.⁻¹) per mole. For elements on the other hand, it appears that the value is between 2 and 3 E.U. per g. atom, a distinct periodicity being evident (E. Kordes, 1927). The latter rule holds approximately for most elements, but the former is of doubtful value; there are, nevertheless, certain generalizations of wide applicability. An entropy of fusion of 2 to 3 E.U. per mole is to be expected for substances which have almost the same rotational, vibrational and structural entropies in the solid and liquid states; this is particularly true for metals and other monatomic elements. The chlorine molecule, however, is diatomic and probably undergoes appreciable rotation in the

* The heat capacities and heat of fusion in this equation refer to the values at the equilibrium vapor pressure of the system.

liquid at its melting point, but not in the solid; the entropy of fusion is 9.6 e.u. per mole. Molecular hydrogen is also diatomic, but since the molecules rotate in the solid, as well as in the liquid, the increase of entropy upon melting is only 2.0 e.u. per mole. In most compounds a large part of the entropy of fusion is due to the change from libration in the solid state (p. 422) to free rotation in the liquid; if, however, there is rotation in the solid, the increase of entropy on melting is small. For example, *tert*-butyl chloride rotates in the solid state and its entropy of fusion is 2.0 e.u. per mole, since there is little rotational, vibrational or structural change on melting. An interesting comparison is provided by cyclohexane and methyl cyclohexane; the former undergoes a transition in the solid state at which free rotation sets in, and the molar entropy of transition is 8.6 e.u., but the entropy of fusion is only 2.3 e.u., making a total of 10.9 e.u. per mole. The methyl compound, however, does not rotate freely below its melting point, and the observed entropy of fusion is 10.9 e.u. per mole.

Various attempts have been made to explain the phenomenon of melting, and to account quantitatively for the entropy, and other changes, associated with fusion. One promising line of study is based on the concept of "order-disorder" which has been used with great success in the study of certain alloys (cf. p. 381). In a solid, especially at low temperatures, there is a state of almost complete order, for each atom vibrates about a definite mean position, and migration from one point to another in the crystal lattice is rare. While a liquid still retains some aspect of order, as will be seen in Chapter VII, there is a considerable degree of disorder, or randomness, in its structure which increases with increasing temperature. The process of melting is thus accompanied by a change from more or less complete order to a state of disorder. Such a transition might be expected to take place gradually, and the observed sharpness of the melting points of solids has been attributed to the fact that it requires less energy for several molecules to move simultaneously out of their places in the crystal lattice than for the same molecules to move singly. Part of the increase of entropy upon fusion is thus due to the greater degree of disorder, while another part is due to the expansion which accompanies the melting of most solids; in addition there will be a considerable gain of entropy if there is at the same time a change from libration to free rotation.¹⁴

POLYMORPHIC CHANGES

Transition Point and Pressure.—It will be shown below (p. 469) that under certain conditions polymorphic substances have a **transition point** at which the two crystalline forms can coexist in equilibrium, whereas above or below this temperature only one or other form, respectively, is stable. The transition temperature is thus analogous to the melting point of a solid, and the Clapeyron-Clausius equation can be deduced for the equilibrium between two crystalline forms of the same substance in the usual manner; the result is

$$\frac{dT}{dP} = \frac{T(V_\beta - V_\alpha)}{L_t} = \frac{T(v_\beta - v_\alpha)}{l_t}, \quad (86)$$

where dT/dP gives the variation of the transition point with pressure, L_t and l_t are the latent heats of transition per mole and per g., respectively,

and V and v are the molar and specific volumes, respectively. The suffixes α and β refer to the two crystalline forms, the former being converted into the latter on raising the temperature above the transition point. For sulfur, which exists in rhombic (α) and monoclinic (β) modifications with a transition temperature of 95.6°C . at atmospheric pressure, the following data may be used:

$$v_\beta - v_\alpha = + 0.0126 \text{ cc.}, \quad l_t = 2.52 \text{ cal./g.}, \quad T = 273.2 + 95.6 = 368.8^\circ;$$

$$\therefore \frac{dT}{dP} = \frac{368.8 \times 0.0126 \times 76 \times 13.59 \times 981}{2.52 \times 4.184 \times 10^7} = 0.045^\circ \text{ per atm.}$$

The transition point between rhombic and monoclinic sulfur should thus be raised 0.045° per atm. increase of external pressure; the experimental value is approximately 0.05° . The quantity dT/dP has a positive value because the specific volume of the β -form is greater than that of the α -form, but if the reverse is the case, as in the transition from β -rhombic to rhombohedral ammonium nitrate at 84.2°C ., the transition point decreases as the external pressure is increased. Mercuric iodide behaves like sulfur at pressures below about 5000 atm., but at higher pressures a reversal occurs.

The latent heat of transition is analogous to other types of latent heat, and a simple application of Hess's law shows that it must be equal to the difference between the heats of sublimation of the two forms at, or very near, the transition temperature. The dependence of the heat of transition on temperature can be represented by an equation similar to (85) where L_t replaces L_1 , and C_α and C_β , the molar heat capacities of the two crystalline forms at constant pressure, replace C_s and C_L respectively. The simple Kirchhoff equation is not applicable (see p. 462).

Determination of Transition Point.—The thermometric method is based on the fact that the change from one crystalline form to another on raising the temperature is accompanied by an absorption of heat, whereas on cooling there is an equivalent liberation of heat. The substance under investigation is heated slowly and regularly, and the temperature noted from time to time; at the transition point there is a break due to the absorption of heat. A similar break is observed on cooling from above the transition point. There is only one temperature, at each pressure, at which the two crystalline forms can coexist, and so on heating or cooling the system must remain at the transition temperature while one form is changing into the other. A modification of the thermometric method is used to study the "recalcescence" phenomenon in iron, which is due to the heat evolution accompanying a change of crystalline form on cooling. By measuring the difference of temperature between the iron and a platinum wire cooling at the same rate, a change in the rate of cooling of the former can be readily detected.

A change in crystalline form is generally accompanied by a change of density; this is utilized in the dilatometer method for determining transition points. In its simplest form the dilatometer consists of a bulb, into which is placed the experimental material, sealed to a capillary tube; the remainder of the bulb contains an inert liquid whose level can be read in the narrow tube.

The bulb is placed in a heating bath and its temperature is raised gradually. At the transition point the change of density causes a marked change of volume, and this can be detected by the rapid movement of the level of the liquid. As the change from one form to another may be slow the experiment should be repeated, the substance now being cooled from above the transition point, and the mean of the two values taken. More accurate results are obtained by putting both crystalline forms into the bulb and then placing the dilatometer into baths at various temperatures. At the transition point the two forms can coexist and there will be no change in volume; above or below this temperature, however, one form will tend to pass into the other, and an expansion or contraction will be detected by the movement of the liquid level.

It will be seen shortly that the two crystalline forms have different vapor pressure-temperature curves which meet at the transition point. Similarly, they have different solubilities which become identical at that temperature. Measurement of vapor pressure or solubility of the two forms at a series of temperatures can consequently be used to determine the transition point. Instead of measuring the solubility directly, other properties of the solution which depend on concentration, e.g., potential of an electrode (see p. 933) or the viscosity, may be employed.

Two allotropic forms of a metal generally have different electrical properties, and consequently the variation of a suitable property with temperature will show a break at the transition point. Electrode potentials and the thermoelectric effect have been employed in this connection.

SOLID-LIQUID-VAPOR EQUILIBRIA

The Triple Point.—The curves OA and OB in Fig. 99 show the variation with temperature of the vapor pressure of liquid and solid forms, respectively, of a given substance: the curve OA , as already seen, extends only as far as the critical temperature, whereas OB continues to the absolute zero, possibly with a change in direction due to a polymorphic transition (see p. 469). The vapor pressure curve OA of the liquid represents the conditions of equilibrium, viz., temperature and pressure, for a system of liquid and vapor, and OB , similarly, indicates the conditions under which solid and vapor are in equilibrium. The two curves meet at O , and hence at this point, known as the **triple point**, the three physical states of the substance, i.e., solid, liquid and vapor, must be able to coexist.

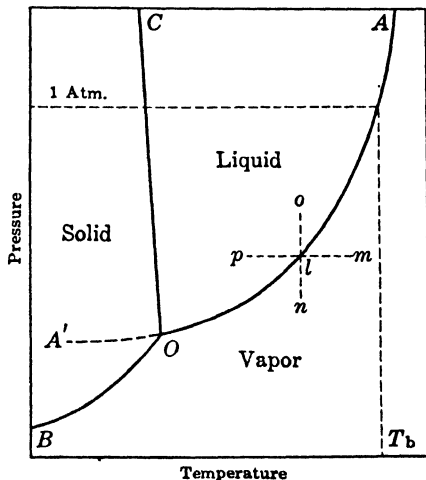


FIG. 99. Solid-liquid-vapor equilibria

That the two vapor pressure curves must cut each other, and in the manner shown, can be deduced qualitatively from the Clapeyron-Clausius equation.

Since the specific volume of liquid or solid is small compared with that of the vapor, the slope of the vapor pressure-temperature curve, i.e., dp/dT , is determined essentially by the latent heat [cf. equation (38)]. From (83) it is evident that the latent heat of sublimation must be greater than that of evaporation, and so the slope of the vapor pressure curve of the solid must always be greater than that for the liquid at the same temperature; this conclusion is in harmony with Fig. 99.

The curve OC indicates the temperatures and pressures at which the solid and liquid forms can be in equilibrium; in other words, it represents the influence of pressure on the melting point, as given by (82). Its slope depends on whether increase of pressure raises or lowers the melting point, i.e., on whether the specific volume of the liquid is greater or less respectively, than that of the solid. In Fig. 99 the curve slopes to the left indicating, as for water and bismuth, that the melting point is lowered by increase of external pressure. It should be noted that the extent of the slope has been greatly exaggerated for the sake of clarity. The curve OC must obviously meet the other curves at the triple point O , where the solid, liquid and vapor are in equilibrium. The conditions under which two of the three phases can coexist is given by each of the curves, and all three forms are in equilibrium where the three curves meet. There is consequently only one point where this is possible, and hence there is only one triple point. The normal melting point of a solid is the temperature at which solid and liquid are in equilibrium at atmospheric pressure; at the triple point, however, the pressure is the equilibrium vapor pressure of the system, and the temperature differs slightly from the melting point. For water, for example, the triple point must be in the vicinity of 0°C ., but the vapor pressure of liquid water and ice is then about 4.6 mm. Since increase of pressure lowers the freezing point of water by 0.0075° per atm., it is apparent that lowering the pressure from 760 mm. (1 atm.) to 4.6 mm., the equilibrium vapor pressure, will *raise* the freezing point by nearly 0.0075° . The triple point of water, found by experiment, is at a temperature of $+0.0098^\circ \text{C}$.* and a pressure of 4.58 mm. of mercury.

The Phase Diagram.—A diagram such as Fig. 99 giving the conditions of equilibrium between various forms, or phases,† of a substance is called a **phase diagram** or **equilibrium diagram**; by its means much interesting information can be obtained. If the pressure of a system consisting of liquid and vapor in equilibrium is maintained constant, and the temperature is increased, it is clear that the whole of the liquid will be eventually converted into vapor; such a change is represented by lm in Fig. 99, so that at m vapor only is present. Similarly, if the temperature is held constant and the pressure diminished, from l to n , complete vaporization of the liquid will occur. If a solid-vapor system, represented by a point on OB , is made to undergo analogous changes, the solid will be converted

* The freezing point of pure water at 1 atm. is 0.0023°C ., so that the calculated triple point is $0.0023^\circ + 0.0075^\circ$, i.e., 0.0098°C .

† The precise definition of the term "phase" is given on p. 474.

into vapor. It will be evident, therefore, that any point in the area lying below the curves OB and OA represents vapor *only*. Returning to the liquid-vapor equilibrium at l , it can be readily deduced that changing the conditions to o , i.e., constant temperature and increase of pressure, or to p , i.e., constant pressure and decrease of temperature, results in the disappearance of vapor, so that liquid alone is present. Application of arguments of the same type to points on OB and to those on OC , shows that only liquid can exist in the area between the curves OA and OC , whereas solid only can be present under conditions of temperature and pressure represented by the area between OB and OC . The single phases that are stable in the different areas are indicated in Fig. 99; two areas meet at a line which gives the conditions under which two phases can coexist, and the three lines meet at the triple point.

Metastable Equilibrium.—It is sometimes possible to continue the vapor pressure of the liquid below the triple point, as shown by the dotted curve OA' ; this results when the liquid is "supercooled," that is, cooled below its freezing point, without solidification occurring. Supercooling of a liquid is one case of the general phenomenon of **suspended transformation**, some illustrations of which were mentioned on p. 432 in relation to the idea of continuity of state. The liquid-vapor system along OA' is then in a condition of **metastable equilibrium**. The term metastable is used to describe a definite equilibrium, which is nevertheless not the most stable equilibrium at the given temperature; a metastable system undergoes spontaneous change on the addition of the stable phase. For example, the addition of a small amount of solid to a liquid-vapor system on the curve OA' will result in immediate solidification of the whole liquid; the temperature and pressure will then be represented by a point on OB . It is seen from Fig. 99 that the metastable system has a higher vapor pressure than the stable one at the same temperature; this is universally true, and indicates the tendency for the former to pass into the latter. The realizable metastable curve OA' is generally short, because a point is eventually reached at which the system undergoes spontaneous change to the stable state.

Sublimation.—A study of the phase diagram provides interesting information concerning sublimation. It can be seen from Fig. 99 that if vapor at a pressure below the triple point is reduced sufficiently in temperature, the vapor will condense directly to the solid form without forming liquid as intermediate stage. This represents the condition under which hoar frost and snow are produced, on the earth's surface and in the upper atmosphere, respectively; if the partial pressure of water vapor in the atmosphere is less than 4.6 mm., then a sudden drop of temperature will result in the direct deposition of solid. For this reason the vapor pressure curve of ice (OB) is often called the **hoar-frost curve**. If the pressure of water vapor is greater than 4.6 mm., liquid water will form on cooling, but hoar frost may be obtained subsequently as the partial pressure becomes less. "Flowers of sulfur" are produced by the condensa-

tion of the vapor direct to solid on a cold surface, so that the pressure is maintained below the triple-point value.

Just as vapor can be converted directly to solid, so solid can be completely vaporized without the intervention of a liquid phase, provided the pressure of the vapor is not allowed to exceed that of the triple point. If the latter is fairly high, then not only is it a simple matter to maintain this condition, but in addition the rate of vaporization of the solid is considerable; it is then possible to purify a solid by the process of sublimation. The triple point of iodine is at 114.15°C ., when the equilibrium pressure is 90.0 mm.; if solid iodine is heated at a temperature below 114° , it will vaporize without melting, and if the vapor is condensed on a cooled surface the pressure will be kept below 90 mm., so that vaporization will continue until the solid has disappeared. On the cold surface condensation takes place from the vapor direct to solid, because the pressure is below that at the triple point. Other substances which can be readily sublimed, and so purified, in the same manner are benzoic acid and naphthalene.

Provided the triple-point pressure is less than 1 atm., fusion of a solid will occur on heating rapidly, so that the vapor pressure can exceed that at the triple point. If iodine is heated in a closed space the pressure in-

creases, and when it attains 90 mm. of mercury the temperature will be 114.15°C . and the solid will melt; further heating results in an increase in the vapor pressure of the liquid which boils at 184°C . at atmospheric pressure. Benzoic acid and naphthalene behave in an analogous manner. Some substances, on the other hand, have triple-point pressures in excess of 1 atm., as shown in Fig. 100; examples are carbon dioxide, gray arsenic and violet phosphorus, for which the pressures are 5.11 atm. (-56.4°C .), 35.8 atm. (818°C .) and 43.1 atm. (595°C .), respectively. These substances cannot be melted by heating at atmos-

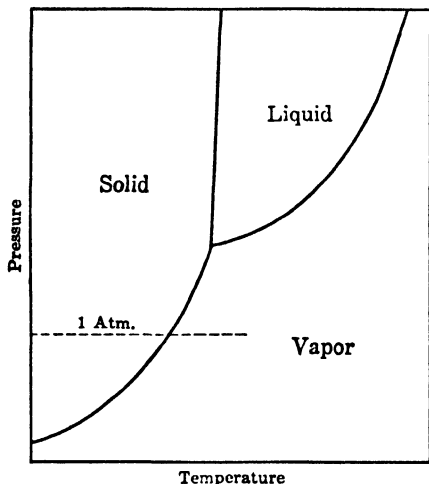


FIG. 100. Phase diagram for sublimation

pheric pressure, neither have they normal boiling points. Liquid carbon dioxide can only be formed if the pressure exceeds 5.11 atm.; at all lower pressures the solid passes directly into gas, a fact which enhances the value of solid carbon dioxide as a refrigerant.

Polymorphism: Enantiotropy and Monotropy.—When a substance exists in polymorphic forms, each modification has its own characteristic vapor pressure curve. Two types of relationship between the curves have

been distinguished, corresponding to the phenomena of **enantiotropy** (Greek: *opposite change*) and **monotropy** (Greek: *one change*).

Two crystalline forms of a substance are said to be enantiotropic, or to exhibit enantiotropy, when each has a definite range of stability, and the change from one form to the other takes place at a definite temperature in *either* direction. This temperature is the transition point (p. 463), and it is the only temperature at which the two solids can coexist in equilibrium at a given pressure; any deviation from this temperature will result in the complete transformation of one form into the other, one being stable above the transition point, and the other below it. Suppose a substance exists in two enantiotropic modifications, α and β , the former being stable below and the latter above the transition temperature; the vapor pressure curves of the solids are then of the type shown in Fig. 101, where AB is the curve for the α -form and BC that of the β -form. At the point B , where the two curves meet, both α - and β -forms are in equilibrium with the vapor, and this is the transition temperature at the vapor pressure of the system; the variation of this temperature with pressure has been already considered. The point C is the triple point or approximately the normal melting point of the β -form, so that CD is the vapor pressure curve of the liquid. It is seen from Fig. 101 that if the α -form of the substance under consideration is heated slowly so that equilibrium is always established, the vapor pressure rises first along AB , at B transformation occurs to the β -form, and then the pressure increases to C where melting takes place. On cooling the liquid, the same changes appear in the reverse order, the β -being converted into the α -form when B is reached. If the α -form is heated rapidly, however, there may be suspended transformation at the transition point; the α -form may then still exist, although in a metastable state, above this temperature and exhibit the vapor pressure curve BE , which is a continuation of AB without a break. The point E is the metastable triple point, approximately equal to the normal melting point of the α -form; from E to C the vapor pressure curve of the liquid is a continuation of CD . Similarly, if the liquid is cooled rapidly, separation of the solid β -form may not take place at C , but supercooling may occur down to E where the α -form commences to crystallize out. The nature of the vapor pressure curves, as in Fig. 101, explains why an enantiotropic substance, e.g., sulfur, melts at different tem-

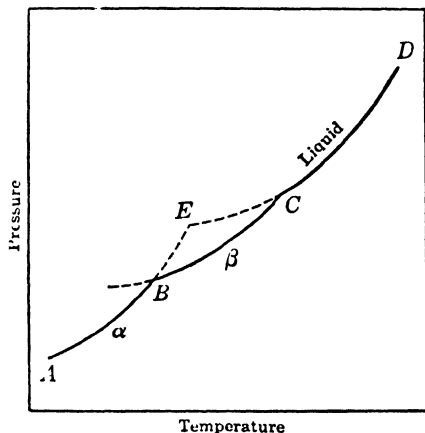


FIG. 101. Vapor pressure of enantiotropic system

peratures according to whether the solid is heated rapidly or slowly. The two melting points are those of the α - and β -forms, respectively. Enantiotropy is a relatively common form of polymorphism; it is exhibited by sulfur, which is considered further below, tin, ammonium nitrate (p. 752), mercuric iodide, carbon tetrachloride and tetrabromide, hexachloroethane and other substances. Transitions accompanying the onset of rotation in the solid state, when there is also a change of crystalline form (p. 423), are invariably enantiotropic in nature.

The phenomenon of monotropy arises when one of the crystalline forms is stable and the other metastable over the whole range of their existence.

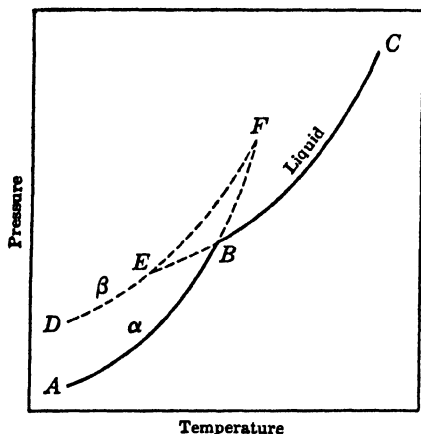


FIG. 102. Vapor pressure of monotropic system

The polymorphic change can then take place in *one direction* only, i.e., from metastable to stable form. The properties of a monotropic system will be clear from the vapor pressure-temperature curves in Fig. 102; AB is that for the α -form, which is always stable, DE is that of the metastable β -form, and BC is the curve for the liquid. The point B is the triple point at which α -solid, liquid and vapor are in equilibrium, and is therefore very close to the normal melting point of the α -form. Similarly, E is the melting point, under its own vapor pressure, of the β -form. If the curves AB and DE are extended they are seen to meet

at F , which is the hypothetical transition temperature; it has, however, no reality, as it is above the melting points of the two modifications and the solids do not then exist. Since the β -form has the higher vapor pressure it is metastable with respect to the α -form, and the former will tend to pass over into the latter, at all temperatures. It is clear from the diagram that under no realizable circumstances can the β -modification be stable with respect to the α -form. Since a stable substance never changes spontaneously into a metastable one, for this would be contrary to the second law of thermodynamics, the transformation from solid α to β is never observed. In order to obtain the metastable β -form it is necessary to melt or vaporize the α -form and to cool the liquid or vapor rapidly. For example, by cooling the liquid quickly, it is possible for the system to pass down CB , through B , and along the metastable curve BE , when the β -crystals will separate at E . Similarly, if the vapor is chilled there may be suspended transformation and instead of stable solid depositing when the conditions are such as to fall on the curve AB , the metastable form will be obtained along DE . It should be noted that although there

is always a tendency for the metastable to change to the stable solid, the rate of transformation may be extremely slow under normal conditions; this is the case with phosphorus, the white form being the metastable one. As is well known the change from the white to the red form occurs very slowly at ordinary temperatures, although it may be accelerated by heating to 260°C ., by light or by the presence of certain catalysts, e.g., iodine. Among other systems exhibiting monotropy, there are silica, iodine monochloride and benzophenone.

Some substances show both enantiotropy and monotropy; for example, although red and white phosphorus are monotropic, there are two forms of the latter which are enantiotropic with respect to each other. The transition point is -76.9°C ., and conversion in either direction is possible; both white forms are, however, metastable and tend to pass into the red crystals. Another instance of combined enantiotropy and monotropy is provided by silica. There are several stable enantiotropic forms, e.g., α - and β -quartz, as well as metastable forms, e.g., α - and β -cristobalite (below 1470°C .), which are enantiotropic with respect to each other; the latter are monotropic with regard to the two modifications of quartz.

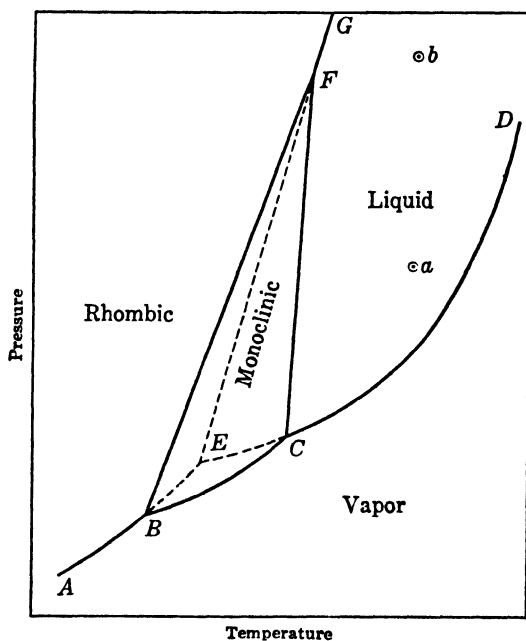


FIG. 103. Phase equilibrium diagram of sulfur (schematic)

The Sulfur System.—The complete phase diagram of sulfur, which exists in the enantiotropic rhombic and monoclinic forms, is shown schematically in Fig. 103; comparison with Fig. 101 shows that AB is the

vapor pressure curve of the rhombic form, stable below the transition point (under its own vapor pressure) *B*, i.e., 95.5°C ., and *BC* is the curve for the monoclinic form, stable above 95.5°C . The melting points of the two forms are at *E* (114.5°C .) and *C* (119.25°C .), respectively; the former can only be realized if the rhombic sulfur is heated rapidly, so that transition to monoclinic does not occur. The curves *BF* and *CF* show the effect of pressure on the transition point and on the melting point of the monoclinic form, respectively; in other words, they give the conditions of equilibrium of rhombic and monoclinic solids (along *BF*) and of monoclinic solid and liquid (along *CF*). The metastable curve *EF*, which continues to the stable curve *FG*, shows the influence of pressure on the melting point of the rhombic form; that is, liquid and rhombic sulfur are in equilibrium along *EFG*. The diagram shows four triple points; the three phases in equilibrium are as follows:

B..... Rhombic, monoclinic and vapor
C..... Monoclinic, liquid and vapor
E..... Rhombic, liquid and vapor
F..... Rhombic, monoclinic and liquid.

The areas in which the four possible phases are stable alone are indicated in Fig. 103; these results can be readily deduced by arguments similar to those on p. 466. It should be noted that whereas the monoclinic form is really stable in the area *BFC*, under certain conditions rhombic can exist in *BEF*, liquid in *CEF*, and vapor in *BEC*; these, however, all represent metastable states.

The triple point *F* (151°C ., and 1290 atm.) is of special interest; * it arises because both the transition point and the melting point of the monoclinic form increase with the pressure, the former more than the latter. The result is that the region of stability of monoclinic sulfur is definitely bounded by the lines *BF*, *CF* and *BC*, and this form cannot exist under any other conditions. Above *F*, therefore, sulfur can be obtained in only one solid modification, namely rhombic crystals. If liquid sulfur represented by a point such as *a* in Fig. 103 is cooled, monoclinic solid will separate when the conditions fall on the curve *FC*, but if the initial temperature and pressure of the liquid are given by *b*, then on cooling, at constant pressure, rhombic crystals must form along *FG*. It is believed that the large crystals of rhombic sulfur found in nature were obtained by direct crystallization from the melt under these conditions.

Forms of Ice.—Other substances, besides sulfur, separate in various crystalline forms according to the pressure; water, for example, will deposit one of five modifications of ice, depending on the pressure. The behavior is different from that of sulfur, however, because only one of the forms, viz., ordinary ice, can exist under normal conditions. There is in addition a sixth form of ice which does not separate from the liquid, but is obtained as a result of poly-

* As the pressure at *F* is so high it has been necessary to exaggerate very considerably the slopes of the curves *BF* and *CF*, so as to bring *F* on to the diagram.

morphic change, at suitable temperatures and pressures, from some of the other modifications. The liquid-solid equilibrium diagram for water derived from the work of G. Tammann (1900 *et seq.*) and of P. W. Bridgman (1912 *et seq.*), is shown in Fig. 104; the point *O* is the ordinary triple point, i.e., $+0.0098^{\circ}$ c. and 4.58 mm., and the curves *OA*, *OB* and *OC* correspond to those in Fig. 99. It has been necessary to exaggerate the position of *O*, and the slopes of *OA* and *OB*, in order that these curves may be shown, for the pressures are low compared with those necessary to illustrate the changes in the solid state; apart from this, Fig. 104 is drawn approximately to scale. The regions of stability of the crystalline forms of ice are marked by Roman numerals; there is some doubt about ice IV and so it has been omitted. It is seen that at increasing pressures

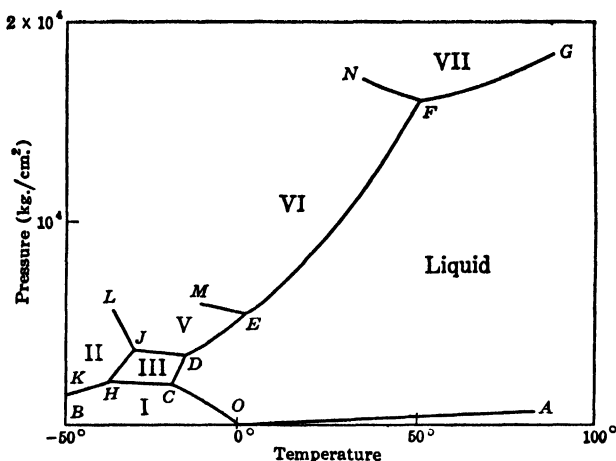


FIG. 104. Phase equilibria of ice

the solids separating from liquid water are ice I, III, V, VI and VII; on the other hand, ice II can only be obtained from other solid forms, e.g., I, III or V, by suitable alteration of temperature or pressure. At the triple points *C*, *D*, *E* and *F* the three phases in equilibrium are two forms of ice and liquid water; at *H* and *J*, however, three solids can coexist in each case. The curves *CH*, *DJ*, *HK*, *JL*, *EM* and *FN* represent two solid phases in equilibrium; in other words, they show the influence of pressure on the transition points. It is possible that if extended to sufficiently low temperatures *OB* and *HK* would meet, giving a condition at which ice I and II and vapor were in equilibrium. Solid III can never coexist with vapor, and it is possible that the same may hold for ice V, VI and VII. It will be observed that at pressures of the order of 20,000 kg. per sq. cm. water freezes at a temperature of about 100° c. to form ice VII.

Heavy ice (deuterium oxide) can exist in a number of solid crystalline forms similar to those of ordinary ice (P. W. Bridgman, 1935).¹⁵

THE PHASE RULE¹⁶

The conditions of equilibrium considered above, based either on equations derived from the second law of thermodynamics or on phase

diagrams obtained by experiment, which are in harmony with the law, can be conveniently summarized in the form of a simple generalization, known as the **phase rule**. This rule was deduced from thermodynamics by J. Willard Gibbs (1876), but its application to physical chemistry is mainly due to H. W. B. Roozeboom (1884 *et seq.*). Before the rule can be stated it will be necessary to define and explain the terms involved.

Phase.—A phase is defined as *any homogeneous and physically distinct part of a system which is separated from other parts of the system by definite bounding surfaces*. Ice, liquid water and water vapor, for example, are three phases; each is physically distinct and homogeneous, and there are definite boundaries between ice and water, between ice and vapor, and between liquid water and vapor. Each crystalline form of ice, however, constitutes a separate phase, since it is clearly marked off from the other forms. In general it can be stated that every solid in a system is an individual phase; each is separated from the others by a definite bounding surface, viz., the outside of the crystal. A solid solution, being perfectly homogeneous, on the other hand, is a single phase no matter how many chemical compounds it may contain. The same applies to a liquid solution; one liquid layer is one phase, whether it consists of a pure substance or a mixture, as long as it is homogeneous. If there are two layers, however, as with ether and water, there are two phases separated by a definite surface. A gas, or mixture of gases, is homogeneous because of the intimate mixing of the molecules; such a system always constitutes only one phase.

Component.—The number of **components** of a system at equilibrium is *the smallest number of independently variable constituents by means of which the composition of each phase present can be expressed, either directly or in the form of a chemical equation.** The water system, for example, consists of one component, viz., H_2O ; each of the phases in equilibrium, i.e., solid, liquid and vapor, may be regarded as being made up of this component only.† It is true that the molecular complexity of water is different in the three phases, but the number of components is not affected. The same applies to acetic acid which consists entirely of double molecules in the solid state, to a great extent in the liquid, and partially in the vapor; the composition of each phase can, however, be expressed in terms of $\text{C}_2\text{H}_4\text{O}_2$, and this is the only component. All the systems discussed in the present chapter consist of one component, but the phase rule is also applied to more complex systems, and some of these will be considered here. For example, in the equilibrium



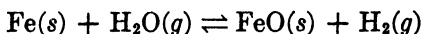
* It should be understood that such equations are restricted to those reactions with respect to which the system is in equilibrium.

† Strictly speaking each isotopic form of water is to be regarded as a separate component, so that ordinary water has, theoretically, two components, viz., H_2O and D_2O ; as long as the ratio of the isotopes remains constant, however, it may be assumed to consist of one component. The form HDO is in any case not a separate component, as it may be expressed in terms of H_2O and D_2O by a chemical equation.

there are three phases, viz., solid CaCO_3 , solid CaO and gaseous CO_2 ; this system has two components, and these may be chosen as *any* two of the substances involved in the equilibrium. If CaO and CO_2 are taken, then the composition of the calcium carbonate phase can be expressed as $x \text{ CaO} + x \text{ CO}_2$ giving, by a chemical equation, $x \text{ CaCO}_3$. The composition of the calcium oxide phase is $y \text{ CaO} + 0 \text{ CO}_2$, and that of the gaseous phase is $0 \text{ CaO} + z \text{ CO}_2$. The compositions of the three phases could be expressed equally by taking CaCO_3 and CaO , or CaCO_3 and CO_2 as the components; in these cases it would have been necessary to use a minus sign in the chemical equation, e.g., $y \text{ CaCO}_3 - y \text{ CO}_2$ is clearly $y \text{ CaO}$. The dissociation of any carbonate, oxide, or similar compound involves two components; the same is true in the case of salt hydrate equilibria, e.g.,



where the simplest components are evidently CuSO_4 and H_2O . In the slightly more complicated equilibrium



it is necessary to choose three components in order that the composition of each of the three phases can be expressed. The compositions of the two solid phases could be given in terms of Fe and O , but these alone are insufficient to define the gaseous phase which is a mixture of hydrogen and water vapor; a third component, viz., H_2 , is necessary.

If ammonium chloride is vaporized into a vacuum, the system consists of one component only, in spite of the dissociation of the vapor into ammonia and hydrogen chloride; the *resultant* composition of the gaseous phase is still represented by NH_4Cl . If, however, an excess of ammonia or of hydrogen chloride is introduced, there are two components; the vapor no longer has the same ultimate composition as the solid.

Degree of Freedom or Variance.—The number of **degrees of freedom** of a system is *the number of variable factors, such as temperature, pressure and concentration, which need to be fixed in order that the condition of a system at equilibrium may be completely defined*. Since the expression "degrees of freedom" has been used in another connection (p. 329), the term **variance** is often employed as an alternative; there is, however, little risk of confusion, although it is common to speak of systems having one, two or three degrees of freedom as being univariant, bivariant or trivariant, respectively. A system consisting of one phase only, e.g., solid, liquid or gaseous, of water has two degrees of freedom, i.e., it is bivariant, for it is evident from Fig. 99 that it is necessary to specify both temperature and pressure to define completely the state of the system. When two phases are in equilibrium, however, the conditions must correspond to a point on one of the lines OA , OB or OC in Fig. 99; only temperature or pressure need be arbitrarily fixed, therefore, in order to define the system, and the latter has one degree of freedom, i.e., it is univariant.

Finally, since three phases can only coexist at a triple point there are then no degrees of freedom, for the fact that the three given phases are in equilibrium automatically fixes the temperature and pressure; such a system is said to be invariant.

The Phase Rule.—Provided the equilibrium between any number of phases is not influenced by gravity, by electrical or magnetic forces or by surface action, and only by temperature, pressure and concentration, then the number of degrees of freedom (F) of the system is related to the number of components (C) and of phases (P) by the phase rule equation

$$F = C - P + 2$$

for any system at equilibrium at a definite temperature and pressure. Consider the one-component system of water; it has been seen above, from the experimental results, that systems of one, two and three phases, respectively, in equilibrium have degrees of freedom of two, one and zero. Since $C = 1$, it follows from the phase rule that $F = 3 - P$; the sum of the number of phases and degrees of freedom should always be three, in agreement with observation.

In the sulfur system four phases can be obtained under normal conditions; according to the phase rule, however, it is impossible for all four to coexist in stable equilibrium. Putting $C = 1$, since the system has only one component, and $P = 4$, it is seen that F would be -1 ; one factor too many has been fixed, and hence three phases is the maximum number which can be in equilibrium. For $P = 3$, $F = 0$ and the system will be invariant; systems consisting of three phases must therefore be represented by fixed points on the pressure-temperature diagram. Writing R for rhombic, M for monoclinic sulfur, L for liquid and V for vapor, it can be deduced, without knowing anything further about the experimental behavior, that there should be four such fixed, i.e., triple, points; viz., for the equilibria $R-M-V$, $M-L-V$, $R-L-V$ and $R-M-L$, there being four ways of combining three out of four different objects. The experimental facts depicted in Fig. 103 bear out the expectation from the phase rule; the four triple points are B , C , E and F , respectively. When two phases are in equilibrium, for which the sulfur system offers six possibilities, viz., $R-V$, $M-V$, $L-V$, $R-M$, $M-L$, $R-L$, there should be only one degree of freedom; such equilibria should, therefore, be represented by curves on the pressure-temperature diagram. Examination of Fig. 103 shows that there are actually six curves, viz., AB , BC , CD , BF , CF and EFG , respectively, corresponding to the theoretically possible pairs of phases. Finally, for one phase alone the phase rule requires two degrees of freedom, i.e., both temperature and pressure; a single phase should occupy an area on the equilibrium diagram, and for sulfur there should be four such areas. This is again in harmony with the experimental facts. It should be pointed out that the phase rule is unable to distinguish between a metastable and a stable system, for the former still represents a state of equilibrium, even if not the most stable.

It will be evident from the two simple cases considered that the phase rule defines correctly, if qualitatively, the conditions of equilibrium between the possible phases in one component systems. Further applications to systems of two or more components will be discussed in Chapter X.

Deduction of the Phase Rule.—In a system of C components distributed between P phases, the composition of each phase is completely defined by $C - 1$ concentration terms, for if the concentrations of all but one of the components are known, that of the last one must be equal to the remainder. Hence for the compositions of the P phases to be defined it is necessary to have $P(C - 1)$ concentration terms; that is to say, the total number of concentration variables is equal to $P(C - 1)$. In addition to the composition, the temperature and pressure of the system, which are the same for all the phases, are variables; hence, assuming that no other forces influence the equilibrium, it follows that

$$\text{Total number of variables} = P(C - 1) + 2.$$

It is now necessary to see how many variables are defined by the fact that the phases are in equilibrium; the difference between the total and this number will give the number of degrees of freedom of the system. Consider the components, designated by 1, 2, 3, etc., in equilibrium between two phases a and b ; the chemical potentials of the various components in the two phases may be represented by $\mu_{1(a)}$, $\mu_{2(a)}$, $\mu_{3(a)}$, etc., and $\mu_{1(b)}$, $\mu_{2(b)}$, $\mu_{3(b)}$, etc. It has been seen (p. 239) that for any *closed* system in equilibrium at a definite temperature and pressure, $\sum \mu dn = 0$; this condition applies to the whole system of P phases, since this represents a closed system. Suppose a small quantity δn_1 moles of component 1 is transferred from phase a to phase b ; then it follows that

$$-\mu_{1(a)}\delta n_1 + \mu_{1(b)}\delta n_1 = 0;$$

$$\therefore \mu_{1(a)} = \mu_{1(b)}. \quad (87)$$

It is seen, therefore, that when two phases are in equilibrium at a definite temperature and pressure the chemical potential of a given component is the same in each phase. This conclusion is applicable to all the components, so that

$$\mu_{2(a)} = \mu_{2(b)}; \quad \mu_{3(a)} = \mu_{3(b)}; \quad \text{etc.}$$

Suppose the system consists of three phases, a , b and c , then the equality of chemical potential for a given component must apply to b and c , as well as to a and b ; hence,

$$\mu_{1(a)} = \mu_{1(b)} = \mu_{1(c)} \quad (88)$$

and two *independent* equations determine the equilibrium between three phases. In general, it can be seen that, for P phases and C components, the fact of the system being in equilibrium gives the set of relationships

$$\begin{array}{ccccccc} \mu_{1(a)} & = & \mu_{1(b)} & = & \mu_{1(c)} & = & \cdots = \mu_{1(P)} \\ \mu_{2(a)} & = & \mu_{2(b)} & = & \mu_{2(c)} & = & \cdots = \mu_{2(P)} \\ \cdots & & \cdots & & \cdots & & \cdots \\ \mu_{C(a)} & = & \mu_{C(b)} & = & \mu_{C(c)} & = & \cdots = \mu_{C(P)} \end{array}$$

which constitutes $C(P - 1)$ independent equations. In this way $C(P - 1)$

variables are automatically fixed, leaving

$$[P(C - 1) + 2] - [C(P - 1)] = C - P + 2$$

undetermined. In order to define the system completely, therefore, this number of variables must be arbitrarily fixed, and hence must be equal to the number of degrees of freedom (F); that is,

$$F = C - P + 2,$$

which is the phase rule, as given above.

REFERENCES

Guggenheim, E. A. "Modern Thermodynamics."

Young, S. "Stoichiometry."

1. Andrews, *Phil. Trans. Roy. Soc.*, **159**, 575 (1869); **166**, 421 (1876); Altschul, *Z. phys. Chem.*, **11**, 577 (1893); Caillaud and Colardeau, *Compt. rend.*, **106**, 1489 (1888); *J. de phys.* **10**, 333 (1891); Bond and Williams, *J. Am. Chem. Soc.*, **53**, 34 (1931); Ipatieff and Monroe, *Ind. Eng. Chem. (Anal.)*, **14**, 171 (1942).
2. Caillaud and Mathias, *Compt. rend.*, **102**, 1202 (1886); **104**, 1563 (1887); S. Young, *Phil. Mag.*, **50**, 291 (1900); H. Eyring, *J. Chem. Phys.*, **4**, 283 (1936); see also, Glasstone, "Theoretical Chemistry," Chap. IX.
3. Thomson, *Proc. Roy. Soc.*, **20**, 1 (1871).
4. Traube, *Z. anorg. Chem.*, **38**, 399 (1904); *Trans. Farad. Soc.*, **34**, 1234 (1938); Teichner, *Ann. Phys.*, **13**, 595 (1904); Hein, *Z. phys. Chem.*, **86**, 385 (1914); Clark, *Chem. Rev.*, **23**, 1 (1938); Maass, *ibid.*, p. 17; Barnes, *ibid.*, p. 29.
5. Baynes, *Nature*, **23**, 186 (1880).
6. MacDougall, *J. Am. Chem. Soc.*, **38**, 528 (1916); **39**, 1229 (1917).
7. Raveau, *Compt. rend.*, **123**, 109 (1896); *J. de physique*, **6**, 432 (1897).
8. Kapitza, *Proc. Roy. Soc.*, **147A**, 189 (1934); Simon, *Phys. Z.*, **34**, 232 (1933); Debye, *Ann. Phys.*, **81**, 1154 (1926); Giaque, *J. Am. Chem. Soc.*, **49**, 1864 (1927); **57**, 1175 (1935); **58**, 1032 (1936); **63**, 405 (1941); Kürti and Simon, *Proc. Roy. Soc.*, **149A**, 152 (1935); see also, Taylor and Glasstone, "Treatise on Physical Chemistry," 3rd. ed. Vol. I, Chap. IV (J. G. Aston).
9. Smith and Menzies, *J. Am. Chem. Soc.*, **32**, 897, 907, 1412 (1910); Ramsay and Young, *Phil. Trans. Roy. Soc.*, **175**, 37 (1884); *J. Chem. Soc.*, **47**, 42 (1885); Baxter, *J. Am. Chem. Soc.*, **29**, 127 (1907); Derby, Gutsche and Daniels, *ibid.*, **36**, 793 (1914); Pearce and Snow, *J. Phys. Chem.*, **31**, 231 (1927); Swann and Mack, *J. Am. Chem. Soc.*, **47**, 2112 (1925); Rodebush and DeVries, *ibid.*, p. 2488; Marshall, *ibid.*, **59**, 1161 (1937); Lamb and Roper, *ibid.*, **62**, 806 (1940); Stuckey and Saylor, p. 2922; Mayer and Wintner, *J. Chem. Phys.*, **6**, 301 (1938); Weber and Kirsch, *Phys. Rev.*, **57**, 1042 (1940).
10. Ramsay and Young, *Phil. Mag.*, **20**, 515 (1885); **21**, 33; **22**, 32, 37 (1886); Carr and Murphy, *J. Am. Chem. Soc.*, **51**, 116 (1929); Roehl, *Ind. Eng. Chem.*, **30**, 1320 (1938).
11. Guldberg, *Z. phys. Chem.*, **5**, 374 (1890); Guye, *Bull. Soc. Chim.*, **4**, 262 (1890).
12. Trouton, *Phil. Mag.*, **18**, 54 (1884); Hildebrand, *J. Am. Chem. Soc.*, **37**, 970 (1915); **40**, 45 (1918); *J. Chem. Phys.*, **7**, 233 (1939); Halford, *ibid.*, **8**, 496 (1940); see also, Glasstone, ref. (2).
13. Ramsay and Young, *Phil. Mag.*, **20**, 515 (1885); **21**, 135 (1886); Crafts, *Ber.*, **20**, 709 (1887).
14. Walden, *Z. Elektrochem.*, **14**, 713 (1908); Kordes, *Z. anorg. Chem.*, **160**, 67 (1927); Eyring, et al., *J. Chem. Phys.*, **5**, 896 (1937); Lennard-Jones and Devonshire, *Proc. Roy. Soc.*, **169A**, 317 (1939); **170A**, 464 (1939); Rice, *J. Chem. Phys.*, **7**, 138, 883 (1939); **9**, 121 (1941); **12**, 1 (1944); see also, Glasstone, ref. (2).
15. Tammann, *Ann. Phys.*, **2**, 1, 424 (1900); *Z. phys. Chem.*, **72**, 609 (1910); Bridgman, *Z. anorg. Chem.*, **77**, 377 (1912); *Phys. Rev.*, **48**, 893 (1935); *J. Chem. Phys.*, **3**, 602 (1935).
16. For references see Chapter X.

CHAPTER VII

THE LIQUID STATE

INTERNAL PRESSURE

Internal Pressure.—The attractive forces between molecules vary inversely as the seventh power of their distance apart (p. 300), and so these forces must be of considerable magnitude when the intermolecular distances are of the order of little more than a molecular diameter. This is the case with liquids, and the powerful forces of cohesion which thus operate are responsible for the main properties of this form of matter. At very short distances of approach repulsive forces become appreciable between the molecules, but these are generally of a smaller order than in solids because the separations are somewhat greater. Nevertheless, it may be supposed that in the interior of a liquid there is a balance between the attractive and repulsive forces, and this gives rise to what is called the **internal pressure**. The quantity $(\partial E/\partial V)_T dV$ is equal to the change of internal energy resulting from an increase of volume, at constant temperature, and this may be regarded as a measure of the work done in overcoming the internal pressure during the volume change in the liquid; calling this pressure P_i ; it is possible to write, at least approximately,

$$P_i dV = \left(\frac{\partial E}{\partial V} \right)_T dV; \quad (1)$$

$$\therefore P_i = \left(\frac{\partial E}{\partial V} \right)_T, \quad (2)$$

so that the internal pressure may be defined by (2), in terms of the influence of volume on the internal energy of the liquid. It has been deduced thermodynamically (p. 236) that for any substance

$$\left(\frac{\partial E}{\partial V} \right)_T = T \left(\frac{\partial P}{\partial T} \right)_V - P, \quad (3)$$

and since the external pressure P is generally small relative to the internal pressure, the latter is sometimes defined by

$$P_i = T \left(\frac{\partial P}{\partial T} \right)_V. \quad (4)$$

The internal pressure can thus be determined experimentally by studying the increase of pressure with temperature.

From the definitions (p. 235) of the coefficients of thermal expansion (α) and of compressibility (β), it follows that $(\partial P/\partial T)_V$ is equal to α/β , so that

by (4),

$$P_i = T\alpha/\beta. \quad (5)$$

This relationship provides a relatively simple method for determining the internal pressure. The van der Waals equation

$$\left(P + \frac{a}{V^2}\right)(V - b) = RT \quad (6)$$

applies to both gaseous and liquid states (p. 440), and differentiation of (6) with respect to temperature at constant volume gives

$$\left(\frac{\partial P}{\partial T}\right)_V = \frac{R}{V - b}; \quad (7)$$

$$\therefore T \left(\frac{\partial P}{\partial T}\right)_V = \frac{RT}{V - b} = P + \frac{a}{V^2}. \quad (8)$$

It follows, therefore, from (2), (3) and (8) that

$$P_i = \frac{a}{V^2}, \quad (9)$$

so that the van der Waals attractive term, a/V^2 , may be taken as a measure of the internal pressure of the liquid; this is not exact, however, because (7) is based on the assumption that the van der Waals equation is correct. Another method of estimating P_i is to assume that the internal molar latent heat of vaporization (p. 444) is a measure of the work done against the internal pressure in vaporizing 1 mole of liquid, occupying a volume V , so that

$$\left(\frac{\partial E}{\partial V}\right)_V V \approx L_i; \quad (10)$$

$$\therefore P_i = \left(\frac{\partial E}{\partial V}\right)_T \approx \frac{L_i}{V}. \quad (11)$$

J. H. Hildebrand (1936) has shown that at ordinary temperatures L_i is approximately equal to $-1400 + 24.5 T_b$ cal., where T_b is the normal boiling point on the absolute scale, and so it is a relatively simple matter to evaluate the internal pressure.

The methods described above do not always give identical results, but they are of a similar order; the values depend to some extent on the external pressure and on the temperature. Some approximate internal pressures, under normal conditions, relative to that of naphthalene, are given in Table 72; the internal pressure of naphthalene is about 3600 atm., so that the forces of cohesion in liquids are evidently very large.¹

TABLE 72. RELATIVE INTERNAL PRESSURES OF LIQUIDS

(Naphthalene = 1.00)			
Hexane	0.56	Ethylene dibromide	1.13
Ethyl acetate	0.73	Carbon disulfide	1.18
Carbon tetrachloride	0.81	Phenol	1.40
Benzene	0.96	Iodine	1.82
Chlorobenzene	0.96	Water	4.55

SURFACE TENSION AND SURFACE ENERGY

Properties of Liquid Surfaces.—A molecule in the interior of a liquid is completely surrounded by other molecules, and so, on the average, it is attracted equally in all directions. On a molecule in the surface, however, there is a resultant attraction inwards, because the number of molecules per unit volume is greater in the bulk of the liquid than in the vapor. As a consequence of this inward pull the surface of a liquid always tends to contract to the smallest possible area; it is for this reason that drops of liquid and bubbles of gas in a liquid become spherical, the surface being then a minimum for the given volume. In order to extend the area of the surface it is obviously necessary to do work to bring the molecules from the bulk of the liquid into the surface against the inward attractive force; the work required to increase the area by 1 sq. cm. is called the **free surface energy**. The tendency for a liquid to contract may be regarded as a consequence of its possession of free energy, since approach to equilibrium is always accompanied by a diminution of free energy (p. 234).

As a result of the tendency to contract, a surface behaves as if it were in a state of tension, and it is possible to ascribe a definite value to this **surface tension**, which is the same at every point and in all directions *along* the surface of the liquid. It is given the symbol γ , and may be defined as the force in dynes acting at right angles to any line of 1 cm. length in the surface. The work done in extending the area of a surface by 1 sq. cm. is equal to the surface tension, which is the force per cm. opposing the increase, multiplied by 1 cm., the distance through which the point of application of the force is moved. It follows, therefore, that the surface energy in ergs per sq. cm. is numerically equal to the surface tension in dynes per cm. It will be noted that both surface tension and surface energy have the same units, viz., mlt^{-2} , and so the quantities are mathematically identical. Although the surface energy is probably to be regarded as the fundamental property of a surface, it is often convenient, for purposes of calculation, to replace it by the surface tension; the equivalence of the two quantities makes this justifiable.

The existence of a surface implies a separation between two media, e.g., liquid and gas, and the surface tension or surface energy is dependent on the nature of both substances. As a general rule the recorded values of the surface tensions of liquids are those for the interface between the liquid and air saturated with the vapor, although somewhat different values are obtained in the absence of air or in the presence of other gases; the surface tension of mercury, for example, is particularly sensitive in this connection.

The work required to enlarge the surface of separation between two immiscible or partially miscible liquids is called the **interfacial surface energy**, and is often expressed as the **interfacial tension** in dynes per cm. The interfacial tension between two liquids is generally less than the larger of the two surface tensions; the mutual attraction across the inter-

face between the molecules of one liquid and those of the other tends to reduce the inward pull of the molecules in the surface by those of its own kind. The tensions at liquid-solid interfaces are of theoretical importance, but very little quantitative information concerning them is yet available.

Contact Angle and Wetting.—Suppose a liquid L rests on the surface of a solid S and comes to equilibrium with the **angle of contact**, measured in the liquid, equal to θ , as shown in Fig. 105; G is the gas, generally air,

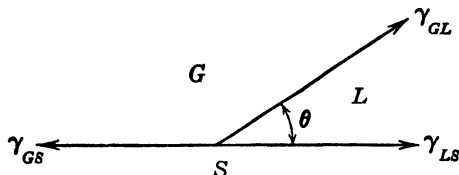


FIG. 105. Gas-liquid-solid interfaces

in contact with the solid and liquid. At equilibrium the forces acting at the interfaces must balance, and assuming that these can be represented by surface tensions acting in the direction of the surfaces, it follows that

$$\gamma_{GS} = \gamma_{LS} + \gamma_{GL} \cos \theta, \quad (12)$$

where γ_{GS} , γ_{LS} and γ_{GL} are the surface energies, or tensions, of the gas-solid, liquid-solid and gas-liquid surfaces, respectively. The contact angle θ thus depends on the three interfacial tensions, but whether it is greater or less than 90° is governed by the relative magnitudes of γ_{GS} and γ_{LS} . If the gas-solid tension (γ_{GS}) is greater than that for the liquid-solid interface (γ_{LS}) then $\cos \theta$ must be positive, and θ is less than 90° , but if the reverse is true then θ must lie between 90° and 180° . In the former instance, e.g., water on glass, the liquid is regarded as wetting the solid, whereas in the latter, e.g., mercury on glass, or water on paraffin wax, the liquid is said not to wet the solid.* The effects on a plane surface and in a tube for the two types of behavior are shown in Fig. 106 *A* and *B*, respectively. It is seen that the surface, or meniscus, of a liquid in a tube is not flat but curved; if the liquid wets the solid then it is concave upward, but if it does not it is concave downward. These phenomena, exhibited by water and mercury, respectively, in glass tubes are very familiar.

The Dupré Equation.—The subject of wetting and of the magnitude of the contact angle may be considered from another point of view. When the surface between L and S in Fig. 105 is diminished by 1 sq. cm., the

* The difference is actually a matter of degree, since complete wetting implies a contact angle of 0° , and absolutely no wetting an angle of 180° ; it is, however, a common convention to say that wetting occurs when θ is less than 90° , but no wetting if it is larger than this value.

interfaces between G and S , and between G and L are both *increased* by 1 sq. cm., and the resultant work done is $\gamma_{GS} + \gamma_{GL} - \gamma_{LS}$. This work is a measure of the energy required to separate the solid from the liquid and is called the **work of adhesion** between solid and liquid; it is given the symbol W_{LS} , thus

$$W_{LS} = \gamma_{GS} + \gamma_{GL} - \gamma_{LS}. \quad (13)$$

This is a form of the equation first deduced by A. Dupré (1869) and generally known by his name.

If the liquid is imagined to be in the form of a column 1 sq. cm. cross section, and the two ends of the column are pulled so that the liquid is divided into two parts, without any lateral contraction, then two new liquid-gas surfaces, each of 1 sq. cm. area, will be formed. The work required is clearly $2\gamma_{LG}$, and since this must be done against the forces of

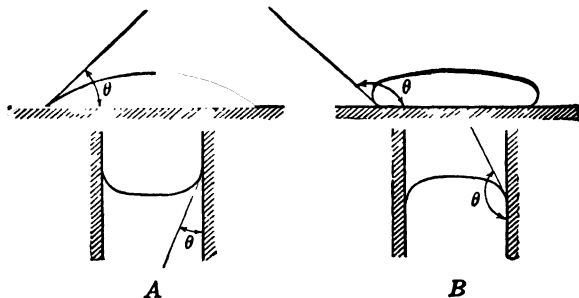


FIG. 106. Contact angle and wetting

cohesion it is called the **work of cohesion** of the liquid. If (12) and (13) are combined it is seen that

$$W_{LS} = \gamma_{LG}(1 + \cos \theta), \quad (14)$$

so that the contact angle θ depends on the relative values of the adhesional work of the solid and liquid, and the cohesive work of the liquid, that is upon the relative attraction of the solid and liquid, and between the molecules of the liquid themselves. Provided the cohesive work is less than half that of adhesion, i.e., the attraction of the liquid for the solid is more than half that for itself, θ is less than 90° and wetting is said to occur. As is to be expected, from general considerations, wetting of a solid by a liquid is therefore the result of a relatively high degree of attraction between the molecules of solid and liquid.

Two Liquids and a Solid.—The equations developed above for gas-solid-liquid interfaces apply equally to the conditions at the interfaces between a solid and two immiscible liquids A and B (Fig. 107). At equi-

librium the analogue of (12) is

$$\gamma_{BS} = \gamma_{AS} + \gamma_{AB} \cos \theta_A, \quad (15)$$

where θ_A is the contact angle of the liquid A with the solid. In Fig. 107 θ_A is shown as less than 90° , so that A is assumed to wet the solid whereas B does not; according to (15) this arises when γ_{BS} is greater than γ_{AS} , so that $\cos \theta_A$ is positive. Of two liquids, therefore, the one which wets a given solid preferentially is the one having the smaller liquid-solid surface tension or surface energy. Application of the Dupré equation (13) shows that the liquid with the larger work of adhesion for the solid will wet the surface of the latter; this is in harmony with expectation. The importance of the relative degree of wetting of a solid by two liquids will be considered in Chapter XIV.

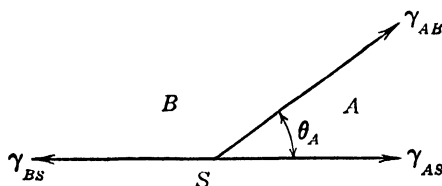


FIG. 107. Interfaces between two liquids and a solid

Spreading of One Liquid on Another.—If a liquid A is to spread spontaneously on another liquid B , in which it does not dissolve, the process must be accompanied by a decrease of free energy. In the spreading, the interface between A and B and that between A and its vapor are increased, whereas that between B and its vapor is decreased, and so for spontaneous spreading it follows that $\gamma_{AB} + \gamma_A - \gamma_B$ must be negative, i.e.,

$$\text{Net work done in spreading 1 sq. cm.} = \gamma_{AB} + \gamma_A - \gamma_B < 0, \quad (16)$$

where γ_A and γ_B are the surface tensions of A and B respectively, against their vapors, and γ_{AB} is the interfacial tension between the two liquids. The form of Dupré's equation for the adhesional work between A and B is

$$W_{AB} = \gamma_A + \gamma_B - \gamma_{AB}, \quad (17)$$

where W_{AB} is a measure of the work required to diminish the surface between A and B by 1 sq. cm., the surfaces of A and vapor, and of B and vapor increasing by this amount at the same time. Combination of (16) and (17) shows that the condition for spreading is

$$W_{AB} > 2\gamma_A; \quad (18)$$

that is to say, the work of adhesion (W_{AB}) between A and B must exceed the work of cohesion ($2\gamma_A$) of the spreading liquid A . The difference

between these two quantities, $W_{AB} - 2\gamma_A$, i.e., $\gamma_B - \gamma_A - \gamma_{AB}$, is called the **spreading coefficient** of *A* on *B* (W. D. Harkins, 1922).

As a result of a comprehensive study of surface tensions of organic compounds Harkins observed that substances having a polar group of the type $-\text{OH}$, $-\text{COOH}$, $-\text{CN}$, or $-\text{NH}_2$, were able to spread on water, whereas hydrocarbons and halogen derivatives did not spread; some of the relevant data are given in Table 73.

TABLE 73. SPREADING OF LIQUIDS ON WATER (DYNES CM.⁻¹ UNITS)

Substance <i>A</i>	Water			Spreading Coefficient
	γ_A	γ_B	γ_{AB}	
Octyl alcohol	27.5	72.5	8.5	36.5
Oleic acid	32.5	72.5	15.5	24.5
Bromoform	41.5	72.5	40.8	- 9.8
Liquid paraffin	31.8	72.5	57.2	- 16.5

The lower homologues of the spreading liquids are soluble in water, and the decrease of solubility with increasing length of the hydrocarbon chain is probably due to the small attraction between the latter and water. This is shown by the low values, e.g., 45 ergs per sq. cm., of the adhesional work between a hydrocarbon and water, as compared with about 90 ergs for an alcohol and water. The latter value is almost independent of the length of the hydrocarbon chain and is consequently to be attributed to the $-\text{OH}$ group, which is thus responsible for the spreading. It will be seen in Chapter XIV that when a long-chain alcohol or acid spreads on water the molecules of the former are apparently oriented in such a manner that the polar group is attached to the water while the hydrocarbon chain points in the opposite direction; this arrangement is in harmony with the relatively high attraction of the polar group for the water molecules compared with that of the hydrocarbon chain.²

Antonoff's Rule.—In developing the conditions for one liquid to spread on another it was assumed that the liquids are not mutually soluble, so that their surface tensions remain unchanged; when each liquid dissolves in the other the tensions are altered, and G. N. Antonoff (1907) suggested the rule that when the two saturated liquid layers are in equilibrium,

$$\gamma_{AB} = \gamma_B - \gamma_A; \quad (19)$$

that is to say, the tension at the interface between the two liquids is equal to the difference of the surface tensions, against vapor or air, of the two mutually saturated solutions. The interfacial tension γ_{AB} is thus less than the larger of the two surface tensions, e.g., γ_B , in agreement with the deductions already made (p. 481).

It follows, according to Antonoff's rule, that when two partially miscible liquids are in equilibrium the spreading coefficient $\gamma_B - \gamma_A - \gamma_{AB}$ is zero. The rule generally applies quite accurately for liquids

having positive "initial" spreading coefficients, that is when one tends to spread on the other in the pure state; * if the coefficient for the pure liquids is negative, the interfacial tension is generally greater than the value given by (19). It is probable that the difference in behavior is due to the fact that the contact angle between the two liquids at equilibrium is almost zero for systems of the former type, but not for the latter. It is only for zero contact angles that $\gamma_B - \gamma_A - \gamma_{AB}$ is a strict measure of the spreading tendency.³

Surface Tension and Pressure Differences.—A consequence of the free energy of a surface is that the pressure on the concave side of a liquid meniscus is greater than that on the convex side, as is evident from the following considerations. If a surface, e.g., that of a bubble of gas in a liquid, is moved parallel to itself in the direction of convexity, there is an increase of area and work must be done; this is only possible if the pressure inside the bubble is greater than that outside. The simplest case to consider is that of a spherical surface, such as a bubble; suppose this to be cut by an imaginary plane into two equal hemispheres. If p dynes per sq. cm. is the excess pressure inside the sphere, there will be a tendency for the two halves to be driven apart by a force of $\pi r^2 p$ dynes, where r is the radius of the sphere. This is counteracted by the surface tension γ dynes per cm., acting along the circle where the two hemispheres meet, which tends to draw them together; this force is $2\pi r\gamma$ dynes, and so for equilibrium

$$\pi r^2 p = 2\pi r\gamma; \quad \therefore p = 2\gamma/r. \quad (20)$$

This result holds not only for a sphere, but for any surface forming part of a sphere. The excess pressure on the concave side is thus dependent on the radius of curvature of the surface and on the surface tension of the liquid.†

Capillary Rise.—It has been seen (p. 482) that in general the surface of a liquid is curved, and so there will be a difference of pressure on the two sides, the magnitude depending on the radius of curvature. If the liquid is in a capillary tube, the radius is small and the pressure difference will be large. Suppose such a tube, in which the liquid forms a meniscus that is concave upwards, is immersed in a large vessel of the liquid wherein the surface is sensibly flat. The pressure immediately below the meniscus, that is, on the convex side, will then be less by $2\gamma/a$, where a is the radius of curvature of the meniscus, than at the same level outside the capillary tube. As a result of this difference of pressure the meniscus is forced up the tube and equilibrium is attained only when the surface inside the capillary is h cm. above that outside; the pressure at the point A

* There are apparently some discrepancies, even for substances with relatively high spreading coefficients, but the reason is not apparent.

† For any curved surface, having r_1 and r_2 as principal radii of curvature, the excess pressure on the concave side is given by $p = \gamma(1/r_1 + 1/r_2)$; for a spherical surface r_1 and r_2 are equal.

is then the same as at *B* in Fig. 108. The hydrostatic pressure of the column of *h* cm. of liquid, i.e., $hg(\rho - \rho')$, is then equal to the difference between the pressure under the curved meniscus and that at the flat surface; hence

$$2\gamma/a = hg(\rho - \rho'), \quad (21)$$

where ρ and ρ' are the densities of liquid, e.g., water, and of the surrounding gas, e.g., air or water vapor, respectively. Since ρ' is very small in comparison with ρ , it follows that

$$\gamma = \frac{1}{2}hg\rho a. \quad (22)$$

If the angle of contact between the liquid and the glass is assumed to be zero, then in a narrow tube the radius of curvature *a* of the meniscus is the same as the radius *r* of the tube, so that

$$\gamma = \frac{1}{2}hgr. \quad (23)$$

In general *a* is equal to $r/\cos \theta$, where θ is the glass-liquid angle of contact, so that $\gamma = \frac{1}{2}hgr/\cos \theta$.

A liquid, such as mercury, which does not wet glass, forms a surface which is convex upwards; hence the pressure just under the meniscus will be greater than at a plane surface. The level of a mercury surface in a capillary tube will consequently be below that of the surrounding liquid; if the angle of contact is assumed to be 180° then (23) will hold, where *h* is the *fall* of level in the tube.

The rise or fall of liquid in a capillary tube is related to the surface tension of the liquid, and consequently phenomena associated with surface tension are frequently considered under the general title of **capillarity** or **capillary action**, even though capillary tubes are not actually involved.

Measurement of Surface Tension: Capillary Rise Method.—The most accurate method for the determination of surface tension of a liquid is based on the measurement of the rise in a capillary tube of known radius. It can be employed over a range of temperature and pressure, the capillary tube being enclosed in a wider one which can be maintained under any desired conditions. Equation (23) is satisfactory for approximate work, but a number of corrections must be made if results of precision are required. In a tube of finite diameter the meniscus is not spherical and the radius of curvature at the bottom of the meniscus is not equal to the radius of the tube; the method of applying the correction for this difference has been discussed and considerably simplified,*

* For tubes of relatively small internal diameter it is often sufficient to add $\frac{1}{3}r$ to the capillary rise; this involves the assumption that the surface is a segment of a sphere, the additional term being an allowance for the hydrostatic pressure of the curved part of the meniscus.

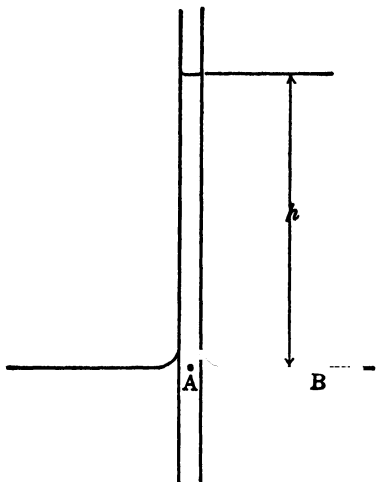


FIG. 108. Capillary rise

with the aid of tables, by S. Sugden (1921). It is necessary in this connection to assume zero contact angle, and this is justifiable if care is taken to clean the capillary tube thoroughly. Another source of error is in the measurement of the capillary rise. This must be determined relative to a flat surface of liquid into which the capillary tube is inserted; to be sufficiently planar the outer surface must have a diameter of at least 4 cm., otherwise the capillary rise is appreciable (T. W. Richards and E. K. Carver, 1921). In the careful work of W. Ramsay and J. Shields (1893), on the variation of surface tension with temperature, the capillary tube was enclosed in a somewhat wider one, but failure to make adequate allowance for capillary rise in the latter introduced an error of as much as 5 per cent in many of their results. The difficulty of obtaining a flat surface of reference has been overcome by measuring the difference of level $h_1 - h_2$ of the menisci in two communicating capillary tubes of different radii; (22) then takes the form

$$\gamma = \frac{1}{2}(h_1 - h_2)g\rho \left/ \left(\frac{1}{a_1} - \frac{1}{a_2} \right) \right., \quad (24)$$

where a_1 and a_2 are the radii of curvature at the bottom of the menisci. In a modified form of the capillary rise method the pressure required to depress the meniscus in the tube to the level of the surrounding liquid is measured (A. Ferguson, 1922); this pressure is, of course, equal to $2\gamma/a$. This procedure has the advantage that it is necessary to determine the radius of the capillary tube at only one point; the corrections for the departure of the meniscus from the spherical, mentioned above, must be applied. Numerous other forms of the capillary rise method for the measurement of surface tension have been described.⁴

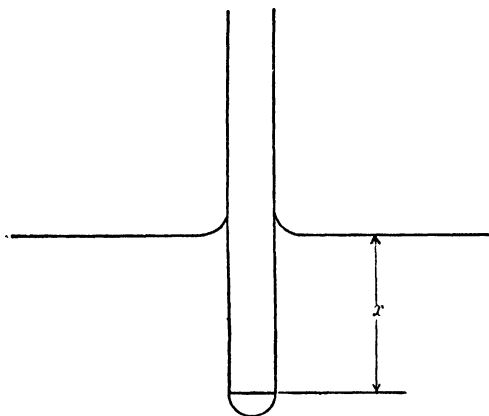


Fig. 109. Maximum bubble pressure

Maximum Bubble Pressure Method.—In order to form a bubble at the end of a tube of internal radius r immersed in a liquid to a depth x (Fig. 109) it is necessary to apply a pressure p , given by

$$p = xg\rho + 2\gamma/b, \quad (25)$$

where the first term is the amount required to overcome the hydrostatic pres-

sure of the liquid * and the second represents the excess pressure on the concave side of the surface of the bubble of radius b . If the internal diameter of the tube is small, the bubble will have the shape of a segment of a sphere, and it can be readily seen that as it increases in size the radius b at first decreases, until the bubble is a hemisphere, and then it will increase. It follows, therefore, from (25) that since the radius is a minimum when the bubble is hemispherical, that is when $b = r$, the pressure must then be a maximum, and

$$p_{\max.} = xgp + 2\gamma/r. \quad (26)$$

In actual practice the pressure of air is maintained at the top of the tube and the bubble becomes unstable and breaks away when the maximum pressure, given by (26), is first exceeded. By measuring this pressure on a suitable manometer the surface tension of the liquid may be evaluated, since x and r can be readily determined. The maximum bubble pressure method has an important advantage over that of the capillary rise; in the latter, impurities tend to collect at the surface and cause serious errors, whereas in the former procedure each bubble produces a fresh surface. Another advantage is that the measured maximum pressure is independent of the angle of contact between the liquid and the tube. The method can be readily adapted for measurement over a range of temperature, provided the pressure can remain in the vicinity of that of the atmosphere.

If accurate results are required it is not justifiable to assume that the bubbles are hemispherical when the pressure is a maximum, for this is only true if the radius of the tube is very small; it is, therefore, not possible to replace b by r for the maximum pressure condition. The method of correction has been discussed by S. Sugden (1922), whose tables have simplified the labor involved. The maximum bubble pressure method is frequently associated with the name of G. Jaeger (1891) who used it for the comparison of surface tensions. In more recent times F. M. Jaeger (1917) employed a tube, adjustable to a known depth below the surface of the liquid, contained in a vessel of at least 4 cm. diameter to avoid capillary effects. An improvement was made by S. Sugden (1924) in which two tubes of different radii, r_1 and r_2 , were employed; they were immersed to the same depth x in the liquid. The maximum pressure was measured for each tube, and if the values are p_1 and p_2 , then from (26)

$$p_1 - p_2 = 2\gamma \left(\frac{1}{r_1} - \frac{1}{r_2} \right). \quad (27)$$

For accurate work r_1 and r_2 must be replaced by the corrected values. It is evident from (27) that the apparatus containing a given pair of tubes can be standardized by means of a liquid of known surface tension, so that it is not necessary to determine the radii of the tubes.⁵

The Drop Weight Method.—The weight of the drop falling from a tube depends on the radius of the latter and on the surface tension of the liquid: the theory of the subject has not been worked out fully, but semi-empirical equations have been developed which may be used for the determination of surface tensions. According to the observations of W. D. Harkins and F. E. Brown (1919), the apparent mass m of a drop falling from a tube of radius r is given by

$$mg = 2\pi r\gamma\phi, \quad (28)$$

* The small additional hydrostatic pressure due to the depth of the bubble itself is neglected.

where $m = m_0(\rho - \rho')/\rho$, m_0 being the mass of the drop in vacuo, and ρ and ρ' are the densities of liquid and surrounding medium, e.g., air, respectively; the quantity ϕ is a correction factor dependent on $r/v^{1/3}$, where v is the volume of the drop. The values of ϕ have been obtained, and tabulated, by Harkins and Brown from experiments on liquids of known surface tension, employing tubes of various dimensions. The weight of a drop of liquid is apparently independent of the angle of contact.

The drop weight method is chiefly used for the comparison of surface tensions; for two liquids 1 and 2, equation (28) gives

$$\gamma_1/\gamma_2 = m_1\phi_2/m_2\phi_1, \quad (29)$$

m_1 and m_2 being the weights of the drops falling from the same tube. If the volumes of the drops are not very different, ϕ_2 and ϕ_1 are almost identical, and so

$$\gamma_1/\gamma_2 = m_1/m_2, \quad (30)$$

thus permitting the surface tensions of the two liquids to be compared; if the value for one is known that of the other can be determined. The weight of the drop may be replaced by the product of the volume v and the density ρ , so that

$$\gamma_1/\gamma_2 = v_1\rho_1/v_2\rho_2. \quad (31)$$

For reasonable accuracy the tip of the tube, where the drops form, should be sharply ground and the drops must form slowly; a device known as a **stalagmometer** (Greek: *drop measurer*), resembling a pipette and ending in a capillary tube which has a flattened tip, is frequently employed for the purpose (I. Traube, 1887). The weight of a definite number of drops is determined, or the number obtained from a given volume of liquid, as it flows between two marks on a stalagmometer, is counted. Another possibility is to measure the volume change when a single drop is slowly forced out from a micrometer syringe which can be read to 10^{-4} cc.⁶

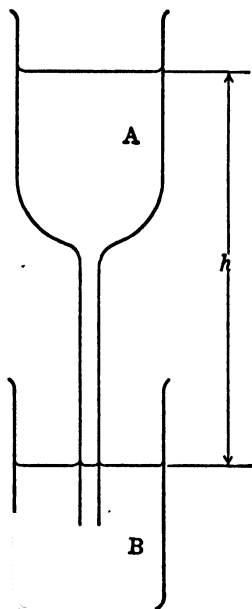


Fig. 110. Interfacial tension

Other Methods.—Various other methods have been described for the determination of surface tension. There is, for example, the “detachment” method, first employed in its simplest form by L. Wilhelmy (1863) and now developed in the form of a commercial instrument known as a “tensiometer” (P. L. du Noüy, 1919). A horizontal ring of platinum wire, attached to one arm of a balance, is immersed in the liquid under examination, and by means of weights, torsion or other devices, the force required to detach the ring from the surface of the liquid is measured; the surface tension is approximately proportional to this force. Mention may also be made of the “sessile” drop and “pendent” drop methods which have been found useful in special cases, e.g., colloidal and similar solutions.⁷

Interfacial Tension.—The capillary rise and drop weight methods are those chiefly used for the measurement of the tension at the interface between two immiscible or partially miscible liquids. The prin-

ciple of the former may be illustrated by Fig. 110; the less dense of the two liquids (*A*) is placed in the upper wide vessel connected with a capillary tube of radius *r* which dips into the heavier liquid (*B*). If the height *h* of *A* is adjusted so that the surface between *A* and *B* is at exactly the same level as that of *B* in the wide vessel, then it can be shown by the method used on p. 487, that an analogous equation,

$$\gamma_{AB} = \frac{1}{2} h g \rho r, \quad (32)$$

is applicable, where ρ is the density of the liquid *A*. The corrections necessary are similar to those described for surface tension. Modifications of the method, involving the use of two capillary tubes, have been described.

In the drop method the usual practice is to measure the volume, in preference to the weight, of the drop; the apparatus generally consists of a form of pipette with the tip bent upwards (F. G. Donnan, 1899), if the less dense of the two liquids is to form the drops, or else up and then down again (W. D. Harkins, 1916) if the heavier is used for this purpose. The number of drops produced by a definite volume of liquid is counted. The interfacial tension is determined by (28), noting that *m* is the *apparent* mass of the drop; thus,

$$m = m_0(\rho - \rho')/\rho, \quad (33)$$

where ρ is the density of the liquid forming the drop, and ρ' that of the other; the value of m_0 is obtained from the volume of the drop by multiplying by its density. The accuracy of the interfacial tension is less the smaller is $\rho - \rho'$, the difference in the densities of the two liquids.⁸

Results.—The highest surface tensions have been observed with molten metals, as shown by the following examples.

	Copper	Gold	Silver	Zinc	Mercury
Temperature	1200°	1200°	1000°	600°	20° c.
Surface Tension	1160	1120	923	770	485 dynes per cm.

Water has a surface tension greater than that of most common liquids, and for many organic substances the values are of the order of 20 or 30 dynes per cm. at ordinary temperatures; the surface tensions of a number of liquids are quoted in Table 74.

TABLE 74. SURFACE TENSIONS OF LIQUIDS AT 20° C. IN DYNES CM.⁻¹

Water	72.8	Toluene	28.4	Acetone	23.7
Nitrobenzene	41.8	Acetic acid	27.6	Methyl alcohol	22.6
Carbon disulfide	33.5	Chloroform	27.1	Ethyl alcohol	22.3
Benzene	28.9	Carbon tetrachloride	26.8	Ethyl ether	17.0

Surface Tension and Temperature.—The surface tensions of liquids generally decrease with increasing temperature; the only exceptions, apart from liquid crystals, are the metals cadmium, iron and copper, but there is as yet apparently no explanation of these abnormalities. An important relationship between surface tension and temperature was deduced by Baron R. von Eötvös (1886) from considerations based on the idea of corresponding states (p. 440); it may also be derived in the following approximate manner. If *v* is the specific volume of the liquid of molecular weight *M*, then *Mv* is the molar volume; if this volume of liquid is spherical, which is the stable form, the surface area is proportional to

$(Mv)^{\frac{1}{3}}$. The product of this molar surface and the surface energy, or surface tension, gives the molar surface energy of the liquid, i.e., $\gamma(Mv)^{\frac{1}{3}}$. This quantity is analogous to the molar volume energy of a gas, i.e., PV , where V is the volume of 1 mole of a gas and P is its pressure. Since PV for an ideal gas is equal to RT , it follows that the derivative of the volume energy with respect to the absolute temperature is a constant; it may be argued by analogy, therefore, that

$$-\frac{d[\gamma(Mv)^{\frac{1}{3}}]}{dT} = k, \quad (34)$$

where k should be a universal constant. The negative sign is introduced because the molar surface energy, like the surface tension, decreases with increasing temperature. This is the Eötvös equation, which implies that the rate of change of molar surface energy with temperature is the same for all liquids independent of temperature. Integration of (34) between the temperatures t_1 and t_2 gives

$$-\frac{\gamma_1(Mv_1)^{\frac{1}{3}} - \gamma_2(Mv_2)^{\frac{1}{3}}}{t_1 - t_2} = k, \quad (35)$$

whereby the Eötvös law may be tested. At the critical temperature, when the meniscus between the liquid and its vapor disappears, the surface tension must be zero, and so it follows from (35) that

$$\gamma(Mv)^{\frac{1}{3}} = k(t_c - t), \quad (36)$$

where t_c is the critical temperature, and γ and v are the surface tension and specific volume, respectively, at any temperature t .

A study of the Eötvös rule was made by W. Ramsay and J. Shields (1893), who found that at temperatures not too near the critical point the molar surface energy could be expressed by a modified form of (36), viz.,

$$\gamma(Mv)^{\frac{1}{3}} = k(t_c - t - 6). \quad (37)$$

Differentiation of (37) again gives (34), and for two temperatures t_1 and t_2 , (35) should still apply. Ramsay and Shields found that the temperature coefficient (k) of molar surface energy was in fact constant and equal to about 2.1, for a number of "normal" substances, e.g., carbon disulfide, carbon tetrachloride, ethyl ether, ethyl acetate, benzene, chlorobenzene, nitrobenzene and benzonitrile. For certain liquids, e.g., water, alcohols and carboxylic acids, the temperature coefficient was not only less than 2.1, but varied with temperature; this abnormal behavior was attributed to the fact, which is undoubtedly true, that these substances are associated in the liquid state (p. 509). If this is the case then, it was argued by Ramsay and Shields, the actual molecular weight of the liquid is not M but xM , where x is the factor of association, that is, the average number of single molecules combined to form a complex molecule; the true tem-

perature coefficient of molar surface energy should then be given by

$$-\frac{d[\gamma(xMv)^{\frac{1}{2}}]}{dT} = k = 2.1, \quad (38)$$

as compared with the observed value k' , based on the molecular weight of the simple molecules; thus $k' = -d[\gamma(Mv)^{\frac{1}{2}}]/dT$. It follows, therefore, that

$$x = (2.1/k')^2, \quad (39)$$

thus permitting the factor of association to be evaluated from the measured temperature coefficient of molar surface energy.

Most of the substances for which Ramsay and Shields found the Eötvös law to apply had molecular weights of about 100, but further measurements by P. Walden (1912) and by F. M. Jaeger (1917) on compounds which are neither associated nor dissociated as liquids, showed that the temperature coefficient of $\gamma(Mv)^{\frac{1}{2}}$ may be as low as 0.56, e.g., for succinic nitrile, or as high as 6, for tristearin. It is probable that ketones have normal molecular weights in the liquid state, but the temperature coefficients k' for acetone, methylpropyl ketone and benzophenone, are 1.8, 2 and 2.9, respectively. Similar differences have been found between nitroethane (1.65) and nitrobenzene (2.2), and between propionitrile (1.5) and benzonitrile (2.1). Metals behave in an abnormal manner, giving small, and sometimes positive, temperature coefficients of molar surface energy, and so also do fused salts; this would imply association, in spite of the fact that the latter are probably dissociated to some extent. It appears that the Eötvös-Ramsay-Shields coefficient is not a constant even for normal liquids, but varies with the molecular weight; Walden suggested the empirical relationship

$$k' = 1.90 + 0.011\Sigma n \sqrt{A}, \quad (40)$$

where A is the atomic weight, and n the number of atoms of each element in a molecule of the compound. This rule is only approximate, for it does not account for normal liquids having values of k' less than 1.9, but it shows that in general the coefficient increases with increasing molecular weight. It is probable, therefore, except perhaps for a few substances, that the Eötvös law is not generally obeyed, and that the method for determining the factor of association used by Ramsay and Shields is quite unreliable. In some instances, for compounds having molecular weights about 100, e.g., aliphatic alcohols, a low value of the temperature coefficient of molar surface energy implies association, although this could not be true for succinic nitrile, acetone, nitroethane and propionitrile. A marked increase of the Eötvös coefficient with increasing temperature is probably the only reliable qualitative indication that the liquid is associated, and certainly no quantitative measure of the extent of association can be obtained at present by any of the equations given above.

A theoretical consideration of the factors affecting the influence of temperature on surface tension suggests that in addition to association, the shape and orientation of the molecules in the surface, as well as their mutual attraction, will have some effect. It is probable that the Eötvös equation will apply only to relatively simple molecules of approximately spherical shape.

The molecular surface energy referred to above is really the *free* energy, and the total molar surface energy can be obtained by use of the Gibbs-Helmholtz equation (p. 232); the free surface energy γ is related to the total surface energy ϵ by the equation

$$\epsilon = \gamma - T \frac{d\gamma}{dT}. \quad (41)$$

The total molar surface energy is equal to $\epsilon(Mv)^{\frac{1}{3}}$, and for normal liquids, including those giving exceptionally high values for the Eötvös factor, this quantity is approximately constant over a range of temperature (G. M. Bennett and A. D. Mitchell, 1913). This can only be true, however, well below the critical point, for as this temperature is approached the total surface energy must gradually decrease to zero.⁹

Surface Tension Relationships.—According to J. D. van der Waals (1894), the surface tension of a liquid at temperature T should be related to the critical temperature T_c by an equation of the form

$$\gamma = \gamma_0(1 - T/T_c)^n, \quad (42)$$

where n is a universal constant, and γ_0 depends on the critical constants of the liquid. The equation has been verified for a number of organic compounds and n is very close to 1.2 (A. Ferguson, 1923; S. Sugden, 1924). The Eötvös equation (36) may be written $\gamma(M/\rho)^{\frac{1}{3}} = k(t_c - t)$, where ρ is the density of the liquid, and M. Katayama (1916) proposed the modified form, in which ρ is replaced by $\rho - \rho'$ where ρ' is the density of the vapor at the same temperature, viz.,

$$\gamma \left(\frac{M}{\rho - \rho'} \right)^{\frac{1}{3}} = k(T_c - T) = kT_c \left(1 - \frac{T}{T_c} \right), \quad (43)$$

the absolute temperature difference $T_c - T$ being, of course, equal to $t_c - t$. Equation (43) holds quite accurately for normal liquids at temperatures much nearer to the critical value than does the original form of Eötvös. By combining (42) and (43), and assuming n to be 1.2, as found experimentally, it follows that

$$\gamma^{\frac{1}{3}}/(\rho - \rho') = C, \quad (44)$$

where C is a constant for each substance (A. Ferguson, 1923). It is of interest that this relationship had been previously proposed on entirely empirical grounds by D. B. Macleod (1923), it is obeyed with considerable accuracy by many organic compounds nearly up to their critical tempera-

TABLE 75. TEST OF MACLEOD'S EQUATION

Benzene				Ethyl ether			
Temp.	γ	$\rho - \rho'$	C	Temp.	γ	$\rho - \rho'$	C
20° c.	28.88	0.8787	2.638	20° c.	17.01	0.7109	2.857
61°	23.61	0.8330	2.647	50°	13.69	0.6713	2.865
120°	16.42	0.7616	2.643	110°	7.00	0.5707	2.865
180°	9.56	0.6657	2.641	140°	4.00	0.4936	2.865
240°	3.47	0.5137	2.657	170°	1.42	0.3785	2.884

tures. The results in Table 75 provide strong confirmation of the reliability of Macleod's equation; the critical temperatures are 288.5° c. for benzene and 193° c. for ethyl ether. For associated liquids the value of C increases slightly as the temperature is raised, but the effect is relatively small; for example, with ethyl alcohol C increases from 2.755 at 20° to 2.845 at 200° c.

It has been shown (R. H. Fowler, 1937) that Macleod's equation can be deduced theoretically, and should hold in the vicinity of the critical point; there is, however, no reason for believing that it should be applicable to other temperatures, and the fact that it does hold is attributed to the cancellation of subsidiary effects. The remarkable constancy of the quantity $\gamma^{\frac{1}{2}}/(\rho - \rho')$ over a large temperature range would, however, appear to imply something more fundamental. The relationship of Macleod's constants to chemical structure will be discussed in Chapter VIII.¹⁰

Vapor Pressure of Curved Surfaces.—A consequence of the existence of surface tension is that the vapor pressure of a curved surface is different from that of a planar surface; the connection can be derived simply by means of thermodynamics. Let p_0 be the vapor pressure over a flat surface and p that for a convex drop of radius r ; if dm g., i.e., dm/M moles, where M is the molecular weight of the vapor, are transferred from the plane liquid to the drop, the increase of free energy (cf. p. 233) is

$$\Delta F = \frac{dm}{M} RT \ln \frac{p}{p_0}, \quad (45)$$

provided the vapor obeys the gas laws. Suppose the addition of dm g. of liquid results in an increase of da in the surface of the drop, then ΔF is also equal to γda , where γ is the surface tension, i.e., the free surface energy per sq. cm., of the liquid. If a is the area of the spherical drop, then $a = 4\pi r^2$, and $da = 8\pi r dr$, where dr is the increase in the radius of the drop accompanying the increase of da in the area. The mass m of the drop is equal to $\frac{4}{3}\pi r^3 \rho$, where ρ is the density of the liquid, and so dm is $4\pi r^2 \rho dr$. It follows, therefore, that

$$da = 2dm/r\rho; \quad (46)$$

$$\therefore \Delta F = \gamma da = 2\gamma dm/r\rho, \quad (47)$$

and equating (45) and (47) gives

$$\ln \frac{p}{p_0} = \frac{2\gamma}{r\rho} \cdot \frac{M}{RT}. \quad (48)$$

It is seen from (48) that the vapor pressure of a drop is greater than that of a flat surface; for most liquids the difference is inappreciable until the radius of the drop is 10^{-6} cm. or less. The calculations given above apply in general to the vapor pressure over any *convex* surface of radius of curvature r . When liquid is transferred to a concave surface, however, the area is decreased and so the free energy change is $-\gamma da$; the final vapor pressure equation is then identical with (48), except that it is preceded by a negative sign. The vapor pressure of water in a capillary tube is thus less than at a flat surface.

It can be shown, in an analogous manner, that the vapor pressure of small particles of solid is greater than that of larger crystals of the same material. From a consideration of the corresponding vapor pressure curves it immediately follows that a finely powdered substance melts at a lower temperature than the massive material, provided a continuous liquid is formed.

Surface Tension of Mixtures.—When two liquids whose surface tensions do not differ very greatly are mixed, the surface tension of the mixture is very roughly a linear function of the concentration. If the surface tensions differ appreciably, however, then addition of quite small amounts of the liquid of lower surface tension results in a marked decrease, e.g., *iso*-amyl alcohol added to water. The reason for this is that any substance capable of lowering surface tension tends to accumulate at the interface (see p. 1205), and hence the latter, where the tension is measured, is relatively much richer in the constituent of lower surface tension than is the bulk of the liquid. Provided the solution is not too dilute, the surface tension of a mixture varies approximately in a linear manner with the logarithm of the concentration of the added compound (cf. p. 1209). A substance which raises the surface tension of a given liquid must have a low concentration in the surface; hence rapid increases of surface tension for the addition of small amounts of material are unknown. Electrolytes generally increase the surface tension of water, but the extent is small, as is to be expected.

The influence of organic compounds in various homologous series on the surface tension of water was studied by I. Traube (1891); he found that, provided the solutions were dilute, the concentrations at which equal lowering of surface tension was observed decreased three-fold for each additional CH_2 group in any given series. This generalization is known as **Traube's rule**.¹¹

VISCOSITY OF LIQUIDS

Viscosity and Fluidity.—Liquids, like gases (p. 279), exhibit the resistance to flow known as viscosity; in general, it is the property which opposes the relative motion of adjacent portions of the liquid and can consequently be regarded as a type of internal friction. As with gases, the **coefficient of viscosity** η is defined as the force per unit area, viz., dynes per sq. cm., required to maintain unit difference of velocity, i.e., 1 cm. per sec., between two parallel layers 1 cm. apart. If two layers dx cm. apart have a difference of velocity du cm. per sec., then the force f acting

per sq. cm. must be

$$f = \eta(du/dx). \quad (49)$$

Liquids having low coefficients of viscosity are said to be "mobile," since they flow readily; on the other hand, when the viscosity is high the liquid is "viscous," and does not flow easily. The reciprocal of the coefficient of viscosity is called the **fluidity** and is given the symbol ϕ , that is $\phi = 1/\eta$; it is a measure of the ease with which a liquid can flow.

When a liquid flows through a narrow tube it is probable that the thin layer of liquid in contact with the walls is stationary; as a result of viscosity, therefore, the next layer will be slowed down to some extent, and this effect will continue, to a diminishing extent, up to the center of the tube. The rate of flow of the liquid, under a given pressure, will obviously be less the smaller the radius of the tube, and the connection between these quantities was first derived by J. L. M. Poiseuille (1844). If a liquid with a coefficient of viscosity η flows with a uniform velocity, at a rate of v cc. in t sec., through a narrow tube of radius r cm. and length l cm., under a driving pressure of P dynes per sq. cm., then

$$\eta = \frac{\pi Pr^4 t}{8vl}. \quad (50)$$

This equation is the basis of methods for determining coefficients of viscosity of liquids, as well as of gases (p. 283). In the derivation of (50) the assumption is made that the flow takes place slowly, and that the liquid leaves the tube with a negligible velocity; since this is not true in practice, a correction must be applied.*

Stokes's Law.—If a solid body is moved through a liquid, the thin layer of the latter in immediate contact with the solid is virtually at rest, just as in the flow of liquid through a tube; as the result of viscosity a viscous drag is exerted on the moving body. In order to maintain a uniform velocity a steady force must, therefore, be applied to overcome the influence of the viscosity of the liquid. It has been found (G. G. Stokes, 1850) that if a small sphere of radius r travels at a velocity u through a fluid, gas or liquid, having a coefficient of viscosity η , the force applied, which just balances that due to viscosity, is given by Stokes's law, viz.,

$$f = 6\pi r\eta u. \quad (51)$$

If the sphere is falling under the influence of gravity, the constant downward force is $\frac{4}{3}\pi r^3(\rho - \rho')g$, where ρ is the density of the sphere and ρ' that of the medium through which it falls. According to (51) the opposing force of viscosity increases with increasing rate of fall of the body, and eventually a constant speed will be attained when the viscous force

* The Poiseuille equation (50) is applicable only to "laminar" or "streamlined" flow; for high rates of flow, especially with tubes of large radius and liquids of low viscosity, the flow may become "turbulent."

is exactly equal to the gravitational pull; that is,

$$\frac{4}{3}\pi r^3(\rho - \rho')g = 6\pi r\eta u; \quad (52)$$

$$\therefore u = \frac{2gr^2(\rho - \rho')}{9\eta}. \quad (53)$$

This form of Stokes's law is applicable to a solid sphere falling through a liquid or gas, or to a drop of liquid falling through a gaseous medium (cf. p. 12).

Measurement of Viscosity.—The direct measurement of viscosity requires the determination of the rate of flow of liquid through a capillary tube of known dimensions, and the application of (50). A few liquids have been studied in this manner, but as the experimental work is tedious use can be made of relatively simple methods whereby the viscosities of two liquids may be compared; if the coefficient of one of them is known that of the other can be evaluated. The method commonly employed involves an instrument known as a **viscosimeter** or **viscometer**, of which several types are available. The principle involved can be easily understood from Fig. 111 which represents a simple viscosimeter (W. Ostwald); it consists essentially of a bulb *A* with a mark above (*x*) and below (*y*), attached to a capillary tube *B* and a storage bulb *C*. A definite volume of liquid is introduced into *C*, sucked into *A*, and the time *t* is observed for the liquid to flow between the marks *x* and *y*; the experiment is repeated with another liquid. The pressure at any instant driving the liquid through the capillary *B* is equal to $h\rho g$, where *h* is the difference in height between the levels of the liquid in the two limbs; this varies during the experiment, but as the initial and final values are the same in every case, it is evident that the applied pressure is proportional to the density ρ of the liquid. It follows, therefore, from (50), since the

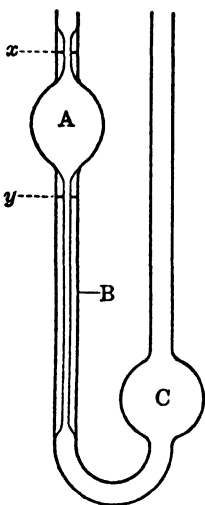


FIG. 111. Ostwald viscometer

same capillary tube is employed, i.e., *r* and *l* are constant, and the same volume *v* of liquid flows through it in each case, that for two liquids 1 and 2,

$$\eta_1/\eta_2 = \rho_1 t_1/\rho_2 t_2, \quad (54)$$

where *t*₁ and *t*₂ are the times of flow. For accurate work, the viscometer designed by E. C. Bingham (1914) is recommended; it depends on the same principle as that of Ostwald, but improvements in design eliminate various sources of error.¹²

Another procedure for the comparison of viscosities, known as the **falling sphere method**, is of particular value for liquids of high viscosity. A small metal sphere is introduced below the surface of a liquid in a tall cylinder, and the time taken for it to fall between two marks is noted. The observation is repeated with the same sphere in another liquid. Since the distance traversed is the same in each case, the velocity of fall is inversely proportional to the

time, and hence from (53) it follows that

$$\eta_1/\eta_2 = t_1(\rho - \rho_1')/t_2(\rho - \rho_2'), \quad (55)$$

where ρ is the density of the metal ball, and ρ_1' and ρ_2' the densities of the liquids. The method is capable of giving accurate results if various corrections are applied; with the aid of (53) it can also be used for the absolute measurement of viscosity. The same principle, in reverse, is employed in the **rising bubble methods**, which has also been applied to the study of viscous liquids. A form of the **oscillating disc method**, described on page 285, has been found especially useful for high temperature measurements, e.g., molten salts and glasses.¹³

Results.—Coefficients of viscosity are generally expressed in c.g.s. units, as indicated in the definition on p. 496; it is readily seen from (50) that the unit is then 1 dyne sec. cm.⁻², and, as stated on p. 279, this is called 1 poise. Since the values in terms of this unit are small, subsidiary units, known as a **millipoise**, equal to 10⁻³ poise, or a **centipoise**, which is 10⁻² poise, are often used. The coefficient of viscosity of water at 20° c. is thus 0.01008 poise or 10.08 millipoises. The viscosities of a number of common liquids are recorded in Table 76.

TABLE 76. VISCOSITIES AT 20° C. IN MILLIPOISES

Ethyl ether	2.33	Methyl alcohol	5.93	Ethyl alcohol	12.0
Acetone	3.29	Benzene	6.47	Acetic acid	12.2
Carbon disulfide	3.68	Chlorobenzene	8.00	Nitrobenzene	20.13
Chloroform	5.63	Carbon tetrachloride	9.68	Glycerol	~10 ⁴

Influence of Temperature.—The effect of temperature on the viscosity of a liquid is strikingly different from that of a gas; whereas in the latter case the coefficient increases with temperature (p. 280), viscosities of liquids invariably decrease markedly as the temperature is raised. Numerous equations connecting viscosity and temperature have been proposed, but those of the exponential type, first derived independently by S. Arrhenius (1912) and J. de Guzmán (1913), viz.,

$$\eta = Ae^{B/RT}, \quad (56)$$

where A and B are constants for the given liquid, are to be preferred for theoretical reasons; it follows from (56) that the plot of $\log \eta$ against $1/T$ will be a straight line. It was thought at one time that the variation of fluidity with temperature would prove more fundamental than that of the viscosity coefficient, but the use of an exponential expression makes the choice immaterial.

An interesting empirical relationship, which appears to be closely obeyed by a number of nonassociated liquids over a range of temperature, was proposed by A. J. Batschinski (1913); thus,

$$\eta = \frac{c}{v - \omega}, \quad (57)$$

where v is the specific volume of the liquid measured at the same temperature as the viscosity; c and ω are constant for each liquid. A rearrangement of (57) gives

$$v = \omega + c/\eta = \omega + c\phi, \quad (58)$$

so that, over a range of temperature, the fluidity should be a linear function of the specific volume; this relationship holds with an accuracy of 1 per cent for over sixty normal liquids, e.g., hydrocarbons, ethers and esters. According to Batschinski the constant ω is related to the critical volume v_c , by the expression $\omega = 0.307v_c$. The extreme values of the numerical factor, for the nineteen substances for which the data were available, were 0.300 to 0.322. According to the van der Waals equation, $\frac{1}{3}v_c$ is equal to the co-volume b , and so ω should be almost identical with the latter. The quantity $v - \omega$ in (57) is consequently equivalent to $v - b$, the "free space," or the volume of the "holes" in the liquid. The viscosity of a liquid is thus approximately inversely proportional to the number of "holes" it contains.¹¹

Viscosity of Mixtures.—Many equations have been proposed to represent the viscosities of mixtures of two liquids, but without an adequate theoretical basis it was not possible to say which corresponded to ideal behavior. Support was at one time obtained for the equation of E. C. Bingham (1906), viz.,

$$\phi = x_A\phi_A + x_B\phi_B, \quad (59)$$

where ϕ is the fluidity of the mixture, and ϕ_A and ϕ_B are the values for the two pure components whose mole fractions in the mixture are x_A and x_B , respectively, but a more satisfactory theoretical relationship appears to be that of J. Kendall (1913),

$$\log \phi = x_A \log \phi_A + x_B \log \phi_B. \quad (60)$$

Mixtures of similar liquids, e.g., benzene and toluene, behave ideally (see Chapter IX) and obey (60). If the constituents of the mixture form complexes in the liquid state, e.g., mixture of amine and phenol, the observed fluidities are less than the calculated values; there is then a considerable increase of viscosity upon mixing. Water and ethanol mixtures exhibit this type of behavior, probably as a result of hydrogen bond formation between the water and alcohol molecules. Mixtures of two dissimilar liquids, e.g., a polar and a nonpolar liquid, such as alcohol and benzene, usually have higher fluidities, and lower viscosities, than would be expected from (60). Such mixtures exhibit what are called "positive deviations" from ideal behavior (p. 711).

Solutions of Long-Chain Molecules.—If η is the viscosity of a solution of concentration c , e.g., in g. per 100 ml., and η_0 is that of the pure solvent, the quantity $(\eta - \eta_0)/\eta_0$ is called the **specific viscosity** of the solution, $\eta_{sp.}$. The value of $\eta_{sp.}/c$ does not vary greatly with concentration, but it is approximately

a linear function of the latter variable, and extrapolation to zero concentration gives what is known as the intrinsic viscosity, $[\eta]$. For solutions of long-chain molecules, such as polymeric compounds, in neutral solvents, e.g., benzene or carbon tetrachloride, the intrinsic viscosity is related to the average molecular weight \bar{M} of the dissolved substance by an expression of the form

$$[\eta] = K\bar{M}^a, \quad (61)$$

where K and a are constants for a given series of long-chain compounds. An equation of this form was suggested empirically by H. Staudinger (1930 *et seq.*) with a equal to unity, but later investigation has shown that for some series of polymers it is less than this value. Further, for substances of molecular weight less than about 5,000 it is necessary to include a small additive term in (61). It appears from theoretical considerations (M. L. Huggins, 1938 *et seq.*) that a in (61) would be unity for a long molecule with its chain kinked in a random manner, but would approach zero if the chain were coiled into a ball. In any event, (61) is of great value for the determination of the average molecular weights of high polymers in solution; it is necessary, however, to determine K and a for the given series of substances by means of two solutions in which the molecular weights are known from other measurements, e.g., osmotic pressure (p. 672).¹⁵

Theory of Viscosity.—Since the influence of temperature on the viscosities of liquids is quite different from the behavior with gases, it has long been suspected that different mechanisms are probably responsible for the flow of liquids and gases. The realization that liquids possess a kind of structure has permitted the development of theories of viscosity which do not involve the transfer of translational energy, from one layer to another, by the interchange of molecules, as is postulated for gases (p. 279). E. N. da C. Andrade (1931, 1934) suggested that the contact between molecules at the extremes of their oscillations (or librations) leads to the transfer of energy; in this way it was found possible to derive an equation of the form of (56). An alternative point of view which has lent itself particularly to quantitative treatment is based on the supposition that each molecule in a liquid may be regarded as occupying, very roughly, an equilibrium position, and before it can pass to another such position in the direction of flow, it must acquire a definite energy of activation.* In general, the higher this energy the greater the viscosity of the liquid, although other factors are also important. According to the Maxwell-Boltzmann distribution law (p. 272) the number of molecules possessing the necessary energy for flow is related to the factor $e^{-E/RT}$, where E is the activation energy per mole. It follows, therefore, that the rate of flow, and hence the fluidity, is proportional to $e^{-E/RT}$; the viscosity will thus depend upon $e^{E/RT}$, as required by (56), in which B may be taken as equivalent to the energy of activation for flow. By means of the "theory of absolute reaction rates" (p. 1098) it is possible to interpret also the factor A in (56); it depends primarily on the molecular weight and molar volume of the liquid, and for many normal, i.e., nonassociated, liquids its value is about 0.5×10^{-3} .

It was suggested by Guzmán that the quantity B in equation (56), i.e., the energy of activation, was related to the latent heat of fusion, but it seems

* The conception of energy of activation is more fully discussed in Chapter XIII.

that there is a closer connection with the heat of vaporization. The energy of activation is regarded as consisting mainly of that required for the formation of a "hole" (p. 432) into which a molecule can pass in its motion from one equilibrium position to the next. Suppose the n molecules forming a liquid are each bound to others by "bonds" adding up to a total energy nU ; to vaporize a single molecule would require energy $\frac{1}{2}U$, since each "bond" is shared between two molecules, provided the rest of the molecules join up so as to leave no hole in the liquid. If a hole is to be left, however, the vaporization of a single molecule requires energy U . The return of a molecule from the gas into a hole prepared for it, in the liquid, would result in the liberation of energy $\frac{1}{2}U$, and so $U - \frac{1}{2}U = \frac{1}{2}U$ is the energy necessary merely to make a hole of molecular size in the liquid, without vaporization of a molecule. It follows, therefore, that the same energy is required to make a hole in the liquid large enough for a molecule, as to evaporate a molecule without leaving a hole. The latter quantity clearly determines the latent heat of vaporization L_v , and so the activation energy E of the viscosity process, which is required for the provision of a suitable hole, should be related to the heat of vaporization. Examination of experimental results shows that this relationship exists, and that liquids fall into a number of well defined groups. For molecules, such as carbon tetrachloride, benzene, nitrogen, argon, etc., which have almost spherical fields of force, L_v/E is about 3, whereas for unsymmetrical molecules, e.g., chloroform, ethyl iodide, ether, long-chain compounds, etc., the ratio is approximately 4. The space required for the molecules to flow into is thus a fraction only of that occupied by a single molecule. For liquid metals the ratio L_v/E is much larger, viz., 8 to 25; this result is in agreement with the assumption, based on the theory of the metallic state (p. 379), that whereas the unit of evaporation is the atom, the unit of flow is the ion; the latter is much smaller and consequently requires less space (H. Eyring, *et al.*, 1936 *et seq.*).

With liquids containing molecules held by hydrogen bonds, E changes markedly with temperature. The viscosity activation energy for these substances consists not only of the fraction, i.e., about one-third, of the energy of vaporization, which is due to the breaking of what may be termed "physical" bonds, but also of the energy of those hydrogen bonds which must be broken when the liquid flows. As the temperature is increased the number of hydrogen bonds in the liquid probably diminishes, because of thermal movements of the molecules, and the energy of activation will decrease. It is incidentally the presence of a network of hydrogen bonds which accounts for the very high viscosities of glycol and glycerol, and for the relatively high values for water and alcohols, compared with other liquids having molecules of the same dimensions. The larger the number of hydroxyl groups in the molecule, the more complex will be the network of hydrogen bonds, and the greater the resistance to flow; the energy required to produce a hole will be large, since work must be done in breaking down part of this system of bonds. In carboxylic acids the hydrogen bonds probably do not extend throughout the whole system, but are probably restricted by the combination of the molecules in pairs (p. 115); it is significant, therefore, that at the same temperature the viscosity of an organic acid is less than that of the corresponding alcohol.*

* This generalization holds for compounds containing more than two carbon atoms.

For long-chain molecules, such as hydrocarbons or polymeric substances, the variation of viscosity with temperature is not represented by

$$\log \eta = a + \frac{b}{T},$$

which is an alternative form of (56), but by

$$\log \eta = a + \frac{b}{T} + cZ^{\frac{1}{2}}, \quad (62)$$

where Z may be taken as the average number of atoms in a chain, and a , b and c are constants. Further, the factor b , which is determined by the activation energy for viscous flow, is found to be constant in any given homologous series and to be independent of the chain length, provided the chain contains more than about thirty atoms. The constant (maximum) value attained by the energy of activation is about 8 kcal., which is much less than would be expected for the long molecules present in the liquid. These results have been accounted for by supposing that such molecules do not flow as a whole but in segments containing 30 or so atoms. The activation energy is determined by the flow of each segment, which will be approximately the same for every member of a given series. The actual viscosity, however, depends on the chain length, as is evident from the presence of the $cZ^{\frac{1}{2}}$ term in (62).¹⁶

Liquid Sulfur.—When sulfur is heated it melts to a mobile liquid, the viscosity of which decreases with temperature in the normal manner; between 160° and 180° c., however, there is a marked increase of viscosity, and the liquid is so thick that it will hardly flow. On further heating the viscosity once more decreases, although relatively slowly, up to the boiling point. If liquid sulfur at 350° c., or more, is rapidly cooled by pouring into cold water, a rubber-like substance, known as plastic sulfur, is obtained; it can be produced as a long continuous thread by running a thin stream of the liquid into water. Plastic sulfur is almost insoluble in carbon disulfide, whereas ordinary sulfur is soluble in this liquid.

The curious behavior of liquid sulfur was attributed by A. Smith (1905 *et seq.*) to an equilibrium between two forms of sulfur, called by him S_λ and S_μ , the former being soluble and the latter insoluble in carbon disulfide. Just above its melting point sulfur was believed to consist of S_λ , but when the temperature is raised it was suggested that increasing amounts of S_μ were formed. By heating liquid sulfur to a particular temperature, cooling rapidly and finding the proportion of soluble and insoluble material in the solid, Smith determined the proportions of S_λ and S_μ in equilibrium in the liquid at various temperatures. The results indicated a rapid increase in the amount of S_μ in the region of 160° to 180° c., and so this substance is presumably responsible for the high viscosity. Plastic sulfur was assumed to consist almost entirely of the S_μ form. A modification of these views was proposed by A. H. W. Aten (1912–14), who considered that a third form of sulfur, called S_τ , very soluble in carbon disulfide, also existed in the liquid phase.

These ideas account for the properties of liquid sulfur in a general way, but there are still some difficulties: for example, if pure sulfur is heated with 2 per cent of sulfur chloride at 180° c., the liquid remains quite mobile. If

this is chilled rapidly, supercooling often occurs and the resulting liquid is completely miscible with carbon disulfide. On the other hand, when the solid separates on rapid cooling, a large proportion is insoluble in carbon disulfide and so presumably consists of S_μ (D. L. Hammick, 1930). This result would suggest that the so-called S_μ does not exist in the liquid, but forms on solidification.

In the sulfur crystal the S_8 molecules are present in the form of puckered rings, and these probably persist in the liquid state above the melting point. As the temperature is raised there is a tendency for the rings to break, and the chains so formed evidently increase in length, and so cause an increase of viscosity. The next homologue to sulfur, viz., selenium, exists in the crystal in long spiral chains continuing throughout the length of the crystal, and the fact that liquid selenium forms a glass readily on cooling suggests that these chains persist to some extent in the fused state. It is not improbable, therefore, that the chains in liquid sulfur will have considerable length. Calculations, based on (62), utilizing accurate determinations of the viscosity of liquid sulfur from 150° to 300° c., indicate that the average chain has a length of the order of 20,000 sulfur atoms at 180° c. when the viscosity is a maximum. With increasing temperature the viscosity and the chain length decrease; the average value of the latter is about 1,000 atoms at 300° c. The energy of activation for the flow of the chains is approximately 9 kcal., while the heat of vaporization is about 2 kcal. per *g. atom*. For a chain molecule E/L_0 is roughly 4, as seen above, and so L_0 per mole should be $4 \times 9 = 36$ kcal.; hence, there must be approximately 36/2, i.e., 18 atoms in each segment of flow.

Upon cooling the liquid, it will be necessary for the long chains first to break and then for eight atoms to join up once more if the S_8 ring present in crystalline sulfur is to be formed; if the chilling is rapid, however, this will not occur and the long chains, perhaps becoming even longer, as in selenium, will become more or less fixed in a fiber-like structure. It is a significant fact, in this connection, that plastic sulfur actually gives an X-ray diffraction pattern typical of fibrous materials. The insolubility of such a substance in carbon disulfide is, of course, not surprising; there is, nevertheless, no reason to believe that the complexity of the sulfur molecule in the plastic solid is the same as in the viscous liquid, as the earlier workers assumed. The failure of liquid sulfur to become viscous in the presence of sulfur chloride must be attributed to the chains remaining relatively short, and these short chains are retained in the supercooled liquid, which is thus miscible with carbon disulfide. On solidification the short chains must either unite to form rings or to give longer chains; it is evident, from the insolubility of the product, that the latter occurs to a large extent. The behavior of liquid sulfur can thus be readily explained without the necessity for postulating definite modifications such as S_λ , S_μ and S_τ .¹⁷

Diffraction of X-Rays by Liquids.—Until a few years ago it was believed that the molecules of a liquid, like those of a gas, were distributed in a random or haphazard manner. The only difference between a liquid and a gas was thought to be that in the former the average distance between two molecules was less, and consequently the attractive forces were greater. In other words, although the liquid state lies between the gaseous and solid states, the general opinion was that it approached much

more to the former than to the latter. An important change of view has, however, occurred mainly as a result of the investigation of liquids by X-rays. In 1916, P. Debye and P. Scherrer showed that when a narrow homogeneous beam of X-rays is passed through a liquid, e.g., benzene, hexane, methyl and ethyl alcohols, and water, a broad, somewhat diffuse, ring or "halo" appeared on the photographic plate at a distance from the position of the incident beam; the fact that this distance was directly related to the wave length of the X-rays established the effect as one of interference. The observations have been confirmed and extended, and it has been shown that a variety of liquids, including liquid elements, hydrocarbons, alcohols and carboxylic acids, give X-ray diffraction patterns consisting of one or two, and sometimes more, broad rings. It was at first considered possible that the diffraction effects were caused by interference of radiations scattered by different parts of the same molecule, such as occurs in gases (p. 591), but since diffraction haloes are given by monatomic liquids, e.g., mercury and argon, it is evident that they are the result of intermolecular, as well as intramolecular, scattering. If the latter alone occurred the intensity of the scattered radiation would be large at very small angles, whereas for liquids the observed scattering under these conditions is small. The production of a definite diffraction pattern, as a consequence of intermolecular scattering, that is, by *different* molecules, implies that a liquid cannot be perfectly "homogeneous" in the same sense as a gas, and that there must be some sort of grouping, or arrangement, of the molecules in a liquid related to that which occurs in the solid state. The idea of an incipient or temporary crystallization is implicit in the views of earlier workers, e.g., W. H. Keesom and J. de Smedt (1922), C. W. Hewlett (1922), and E. D. Eastman (1924), and more recent studies have served to confirm the view that liquids are not entirely free from structure.

Experimental Methods.—The X-ray diffraction of liquids is generally studied by passing a thin beam of monochromatic X-rays through the liquid contained in a thin glass or celluloid tube, or by reflection from the surface of the liquid; the rays are then either allowed to enter an ionization chamber and the strength of the current measured for different scattering angles, or they may fall on a photographic plate from which, after development, a curve giving the scattering intensities at different angles is obtained by a microphotometer device. Two methods have been used to interpret the results; in one, mainly employed by G. W. Stewart (1927 *et seq.*), it is assumed that the Bragg equation (p. 359), viz., $n\lambda = 2d \sin \theta$, is applicable. The angles for which appreciable scattering occurs are inserted, and since the wave length of the X-rays is known the equation can be solved for d ; the values so obtained are assumed to represent the distances apart of important scattering centers in the liquid. The interpretation of the results will be considered shortly.

This method of analyzing diffraction patterns given by liquids is probably oversimplified, and an alternative procedure, the **radial distribution method**, has been developed from the work of P. Zernike and J. A. Prins (1927), and of P. Debye (1925) and P. Debye and H. Menke (1930). The intensity of diffrac-

tion I at different scattering angles, of a liquid consisting of *one kind of atom only*, is given by the expression

$$I = Pnf^2 \left(1 + \int_0^\infty 4\pi r^2 [g(r) - \rho] \frac{\sin sr}{sr} dr \right), \quad (61)$$

where P is the polarization factor (p. 361), f the scattering factor for X-rays of the atoms present, and n their number; the quantity $4\pi r^2 g(r) dr$, involving the "distribution function" $g(r)$, gives the number of atoms lying between spheres of radius r and $r + dr$ surrounding any atom, and ρ is the average density of the liquid in atoms per cc. The symbol s stands for $(4\pi \sin \theta/2)/\lambda$, where θ is the scattering angle and λ is the wave length of the X-rays. The use of the Fourier integral permits the conversion of (61) into

$$4\pi r^2 [g(r) - \rho] = \frac{2r}{\pi} \int_0^\infty s \phi(s) \sin s r dr, \quad (62)$$

where

$$\phi(s) = (I/nP - f^2)/f^2. \quad (63)$$

By means of (62), and the measured intensities of X-ray diffraction at various angles, it is possible by graphical integration to evaluate the quantity $4\pi r^2 [g(r) - \rho]$; in this expression $4\pi r^2 g(r)$ represents the total number of atoms at a radial distance of r from a given atom, and $4\pi r^2 \rho$ is the average value, so that the difference is the number of atoms in *excess of the average* at this distance. The integral in (62) can be evaluated numerically or by means of a harmonic analyzer, making use of the experimental intensities of the X-ray diffraction pattern. The plot of the results against r gives a curve with a series of peaks corresponding to the interatomic distances at which the atomic density exceeds the average; the area under each peak indicates the number of atoms surrounding a given atom at this distance. The radial distribution method is strictly applicable to liquids containing only one important scattering element, or to those in which the scattering powers of the elements, i.e., the atomic numbers, do not differ greatly.¹⁸

Results: Elements.—Application of the Bragg equation to the diffraction patterns of a number of liquid elements gives values of d which are very similar to those observed for the same substances in the solid state. For example, the position of the diffraction ring for liquid helium corresponds to an approximate distance of 3.22 Å. between scattering centers in the liquid, whereas the strongest line in the X-ray spectrum of the solid indicates a spacing of 3.13 Å. Similar results have been obtained for the alkali metals, where the apparent spacing in the liquid is close to that of the (110) planes in the solid, as shown by the results in Table 77

TABLE 77. APPARENT SPACING OF SCATTERING CENTERS IN LIQUID ALKALI METALS

	Sodium	Potassium	Rubidium	Cesium
Spacing in Liquid	3.01	3.87	4.09	4.51 Å.
(110) Spacing in Solid	3.04	3.68	3.97	4.28 Å.

(J. T. Randall and H. P. Rooksby, 1932). The agreement is not so good for bismuth, where the spacings are 2.95 Å. in the liquid and 3.22 Å. for the solid; it is significant that on melting bismuth exhibits a marked increase of electrical conductivity and a decrease in diamagnetism. The results given above suggest that, except for bismuth and probably antimony which has similar physical properties, there is a tendency for the atoms in a liquid to be arranged in a manner having some relationship to that in the solid. Application of the radial distribution method indicates that in liquid sodium (B. E. Warren, 1936), in liquid mercury (J. H. Hildebrand and J. A. Campbell, 1943), and in other cases, the packing of the first layer of atoms is similar to that in the solid; the interatomic distances may, however, be slightly greater in the liquid state. With the second layer the regularity which exists in the solid is hardly apparent in the liquid; the ordered arrangement of the latter is thus "short range" in nature.

Water.—The crystal structure of ice, described on p. 394, consists of oxygen atoms, or ions, each surrounded by four others; it is probable that the hydrogen atoms are situated approximately between two oxygen atoms, which are thus effectively connected by hydrogen bonds. The whole crystal is then virtually one large molecule. The X-ray diffraction patterns of ice and water have some marked similarities, the positions of the diffraction peaks for water corresponding closely to the strongest lines given by ice, and so it is probable that there are similarities in the structures of liquid and solid forms of water. Strong confirmation of this conclusion, based on earlier observations, has been provided by the work of J. Morgan and B. E. Warren (1938), who used the radial distribution method to interpret the diffraction patterns. It appears that at low temperatures there is a tendency for each oxygen atom in liquid water to be surrounded on the average by four others, as in ice, and the distance between two adjacent atoms is 2.90 Å., compared with 2.76 Å. in the solid. The first detailed consideration of the structure of water was made by J. D. Bernal and R. H. Fowler (1933), who suggested that the change from ice to liquid water is analogous to that from the tridymite to the quartz forms of silica; the crystal coordination number is four in each case, but the arrangement of the groups is slightly different. It was considered that a quartz-like structure predominated in liquid water, thus accounting for the increase of density on melting, for the quartz structure is more closely packed than that of tridymite. Morgan and Warren, on the other hand, attribute the higher density of liquid water to the fact that ice has an unusually open structure and that the breaking down of the rigid tetrahedral arrangement of molecules in the solid permits the filling in of the vacant spaces. This effect is opposed by the increase in the intermolecular distances, as shown by the greater separation of the oxygen atoms, and so the temperature of maximum density can be explained.

It is generally agreed that liquid water has a 4-coordinated "pseudo-crystalline" structure, but it differs from a true crystal in the respect that

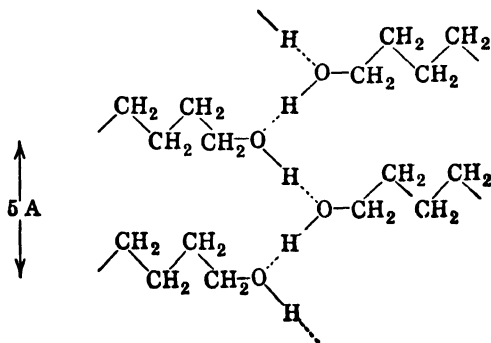
the hydrogen bonds are being continually broken and remade, and that molecules are being interchanged as a result of thermal motions. Although there is a tendency for each molecule to be surrounded tetrahedrally by four others, it is probable that *at any instant* each oxygen atom is joined by hydrogen bonds to only two or three others at ordinary temperatures; at higher temperatures the average number is probably reduced to some extent. These conclusions emphasize the impossibility of ascribing a definite degree of association to water (cf. p. 115), for in a sense the whole liquid, like the crystal of ice, can be regarded as one large molecule, although it is certain that the bonding is much less rigid in the liquid. The unique properties of liquid water must be attributed to the tendency to tetrahedral coordination in the liquid phase; the same structure does not occur in liquid hydrogen sulfide, however, because the hydrogen bonds joining sulfur atoms are much weaker than those involving oxygen.

Organic Compounds: Hydrocarbons.—The normal paraffins from C_6H_{12} to $C_{15}H_{32}$ all give an X-ray diffraction band in almost the same position (G. W. Stewart, 1928); if the Bragg equation is assumed to hold, then these substances have a spacing of 4.63 Å. in the liquid state. It has been suggested that in any small portion of liquid the paraffin chains are approximately parallel and that their separation is about 4.6 Å., independent of their length. It may be significant that $(4.6 \times 10^{-8})^2$ gives 21×10^{-16} sq. cm., which is almost identical with the cross section of a hydrocarbon chain found from a study of surface films (p. 1212). The opinion was expressed by Stewart that the cross sections of the hydrocarbon chains in the liquid must be considered as occupying the corners of a square, but B. E. Warren (1933) has proposed a hexagonal type of packing which is more in conformity with the structure of the crystals of the higher hydrocarbons, $C_{21}H_{44}$ to $C_{29}H_{60}$, just below their melting points. By assuming the centers of the chains to be 5 Å. apart, and taking all possible scattering between atoms in one chain and those in six surrounding chains into account, Warren derived a distribution function which led to a curve for diffraction intensities at different angles very similar to that obtained by experiment. The main point, however, on which there is no disagreement, is that there is a more or less parallel arrangement of hydrocarbon chains in the liquid.

Liquid benzene, the subject of the earliest investigations in the X-ray diffraction of liquids, also gives a single diffraction maximum which, by Bragg's law, implies an apparent distance of 4.7 Å. between scattering centers. The effective thickness of the benzene ring in the solid is 4.6 Å. and so it has been suggested (G. W. Stewart, 1929) that in the liquid there is a definite arrangement in the direction of the thickness of the ring similar to that existing in the solid. The distribution function method, however, leads to a different conclusion (S. Katzoff, 1934; W. C. Pierce, 1937). It is believed that there is a tendency for each molecule of benzene in the liquid to be surrounded by six others in the plane of the ring, with

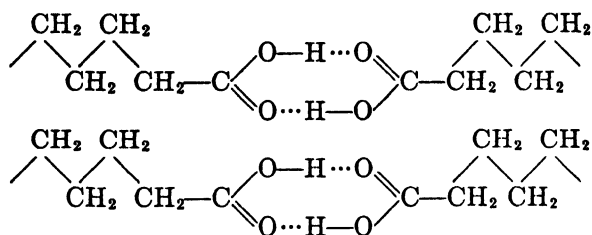
a mean separation of 6.7 Å., and that there are parallel planes of molecules about 3.7 Å. apart. The structure derived in this manner is very similar to that known to exist in the crystal of hexamethylbenzene. It seems, therefore, that there is some regularity of structure in liquid benzene which may be similar to that in the crystal, only of course of a much less rigid nature, as was first suggested by E. D. Eastman (1924). Liquid cyclohexane also gives rise to an X-ray diffraction ring; the intensity peak is sharper than for benzene indicating a more orderly structure (H. K. Ward, 1934).

Alcohols and Acids.—The diffraction patterns of normal primary alcohols and carboxylic acids consist of two peaks; one, corresponding to that observed for paraffins, does not change its position with increasing chain length. The other appears at smaller diffraction angles and is therefore produced by scattering centers at a greater distance apart than those responsible for the first diffraction peak; the distance increases steadily with the length of the hydrocarbon chain. The two peaks are apparently associated with (a) the distance apart of approximately parallel hydrocarbon chains, which is nearly the same in hydrocarbons, acids and alcohols, and (b) the chain length, which increases by about 1.25 Å. for each additional CH_2 group. The appearance of the inner diffraction ring or peak cannot be attributed merely to the presence of the end groups, or oxygen atoms, in the alcohols and acids, for esters do not give this effect. It is probably to be explained by molecular association, involving hydrogen bonds, which results in the grouping of oxygen atoms near a plane passing through the center of the associated molecules at right angles to the chain length (W. C. Pierce, 1938). The scattering from these oxygen layers will produce the inner diffraction ring, its position varying with the number of carbon atoms in the chain. It is significant that mercaptans do not give rise to the inner ring, for in these compounds hydrogen bond formation does not occur readily. On the basis of these views the arrangement of molecules in a liquid alcohol would be approximately as indicated below.



The hydrogen bonds are represented by dotted lines, implying that the hydrogen atom is closer to one oxygen atom than to the other, as in ice (p. 394). The O—H—O distance for hydrogen bonds (see p. 393) is normally 2.5 to 2.8 Å., so that the chains should be about 5 Å. apart, in agreement with the position of the second diffraction peak. The structure depicted is, in general, consistent with the careful analysis, involving the distribution function, of the X-ray diffraction pattern of methyl alcohol. The results suggest that every hydrogen atom of an OH group tends to link two oxygen atoms at a distance of approximately 2.7 Å. apart. The X-ray data for methyl alcohol indicate that each methyl group is surrounded by six others, and it is possible that the hydrocarbon chains in the higher alcohols may be arranged in a similar manner, as postulated by Warren for the hydrocarbons themselves. It is perhaps unnecessary to emphasize that the arrangement shown above is not to be regarded as rigid by any means; there will probably be a continual breaking and re-making of bonds, as described for water, and the chains will not be in one plane. There is some tendency toward a definite arrangement of molecules in the liquid, but it is much less uniform and continuous than in the solid state.

The association in carboxylic acids is of a different type from that in alcohols (p. 115), for in the former definite dimers exist; it is probably the clustering of these in approximately parallel chains, with the carboxyl groups of adjacent molecules lying close together, which leads to the production of a diffraction pattern somewhat similar to that of the alcohols. The idealized arrangement of molecules in the liquid form of the acid may be represented in the following simplified manner.¹⁹



The Structure of Liquids.—The readiness with which liquids flow and diffuse suggests a random molecular structure, analogous to that of a gas; on the other hand, the cohesion and compactness exhibited by liquids indicates a structure resembling that of a solid. Because of the relatively high density of substances in the liquid state, it is clear that each molecule must be simultaneously in the vicinity of many others; since the motion of any molecule is determined by the forces of interaction with the surrounding molecules, it appears inevitable that a liquid should have some kind of structure. The molecular arrangement cannot be completely random as in a gas, until the temperature approaches the critical point.

Near the melting point the density of a liquid is not greatly different from that of the solid. As seen in Chapter V, the packing of atoms (or molecules) in the solid state is usually the most efficient possible. It follows, therefore, that there cannot be a great deal of freedom of movement, which might tend to destroy order, even in the liquid state at low temperatures. As the temperature is raised, however, and the volume increases, there will be an increased randomness of distribution.

The results of X-ray diffraction measurements indicate a certain degree of regularity in the liquid state, as noted above, and the essential differences between the solid, liquid and gaseous states may be considered to be as follows. In a solid there are both "short range" and "long range" order, leading to complete regularity throughout the crystal; in a liquid, the "short range" order still persists to a great extent but the "long range" order is no longer present. Finally, in the gaseous state, there is neither type of order, especially at temperatures well above the critical point. As indicated on p. 463, the phenomenon of melting is to be ascribed to the introduction of an appreciable degree of disorder, and this is largely due to the disappearance of the "long range" order.

As regards the details of liquid structure, two points of view have been put forward, both of which are compatible with the X-ray observations. It is not yet possible to decide between them, and it may be that each is satisfactory under particular conditions. The theory favored by J. A. Prins, J. D. Bernal, and others, attributes to liquids what has been called a "quasi-crystalline" or "pseudo-crystalline" structure. It is supposed that the *average* distribution of molecules about a particular molecule is the same at any instant; the structure differs from that of a crystal, however, because of thermal movements in the liquid. If a number of instantaneous photographs could be taken, the molecular distribution about any molecule in the liquid would appear to be random and homogeneous, but if the pictures were all put together the statistical average would show a definite structure analogous to that of a solid. The other theory, developed by G. W. Stewart (1927), attributes a "microcrystalline" structure to a liquid; according to this view an instantaneous picture would show a large number of "cybotactic groups" (Greek: *space arrangement*), in each of which the molecules, a hundred or more, are arranged in a relatively orderly manner, whereas in the regions between the distribution would be random. The groups are not to be regarded as permanent or as having sharp boundaries, for the movement of the molecules must result in a continual shifting of position and exchange of the members of the cybotactic groups with those in the space between. In view of this continual change, the average of a series of instantaneous photographs would indicate a random arrangement of the molecules.

It is possible that in liquids consisting of long chain molecules, with which the work of Stewart has been particularly concerned, cybotactic groups will exist because of the great length of the molecules; these probably tend to be arranged parallel to each other, especially if they contain

—OH or —COOH radicals. In liquids consisting of more symmetrical molecules, however, it is not so certain that such groups exist. It must be admitted that X-ray methods are not yet sufficiently refined to permit of a distinction between the two points of view described above, and the choice must, for the present, be one of personal preference.²⁰

Theory of the Liquid State.—The thermodynamic properties, such as pressure, energy, heat capacity and entropy, of gases, particularly ideal gases, have been derived from theoretical considerations in quite an adequate manner, and the same may be said of solids, with especial reference to the Debye treatment (p. 418). In the former case the forces between the molecules are regarded as negligible in an ideal gas, and allowance can be made for the molecular interaction in real gases. The problem is here relatively simple, because at pressures that are not too high the interactions involving more than two molecules at any instant may be neglected. In the study of solids, too, the fact that all the molecules occupy fixed equilibrium positions has simplified the problem. With liquids, however, there are certain difficulties which are not easily overcome. The number of different molecular configurations is so large that the mathematical treatment becomes virtually impossible. Several methods have been proposed to avoid these complexities. In that of J. Mayer (1937 *et seq.*) the system is supposed to consist of independent clusters of one, two, three, etc., interacting molecules; this has led to a very satisfactory treatment of condensation and the critical state, but it has thrown little light on the properties of liquids. From the latter standpoint, more satisfactory, although not exact, results have been obtained by the use of a model in which a molecule moves in a "cell" or "cage" made for it by the surrounding molecules. The potential field within the cell may be regarded as spherically symmetrical (J. Lennard-Jones and A. F. Devonshire, 1937 *et seq.*) or it may be assumed that there is a region of virtually uniform potential within the cell (H. Eyring, *et al.*, 1937 *et seq.*). The volume in which the center of a molecule can move in the uniform potential field is called the **free volume**, and this can be related to various thermodynamic properties of the liquid. Another approach to the theory of the liquid state is based on the order-disorder concept, to which reference has already been made; the introduction of "holes" in a liquid makes it possible for the molecules to occupy either the equilibrium positions, as in a solid, or the holes. In this way a degree of disorder is introduced into the system. The fact that several different models have been suggested, none of which is completely satisfactory, indicates the difficulties associated with the development of an adequate theory of the liquid state.²¹

Liquid Helium.—The remarkable fact of the existence of liquid helium in two forms is worthy of mention, although the explanation of the phenomenon is not simple (F. London, 1939). The density, dielectric constant, specific heat and viscosity undergo a marked change at 2.19° K., under its own vapor pressure, although the helium is in the form of a normal liquid before and after the change. The temperature at which the transition from liquid helium I to helium II occurs is known as the **λ -point**; it decreases with increasing pressure, indicating that there is a decrease of density in passing from liquid I to II at the λ -point. This conclusion is in harmony with the fact that helium II gives a more diffuse X-ray pattern than the other liquid form, a result which implies a lower coordination number in the former.

VISCOSITY OF LIQUIDS

A further peculiarity arising in the helium system is that the solid cannot be obtained merely by lowering the temperature of the liquid. It is necessary to increase the pressure at the same time until it is considerably greater than the vapor pressure, so that there is no triple point at which vapor, liquid and solid helium can coexist. The phase equilibria are represented diagrammatically in Fig. 112; the curves *OA* and *OB*, which are greatly exaggerated in height, rela-

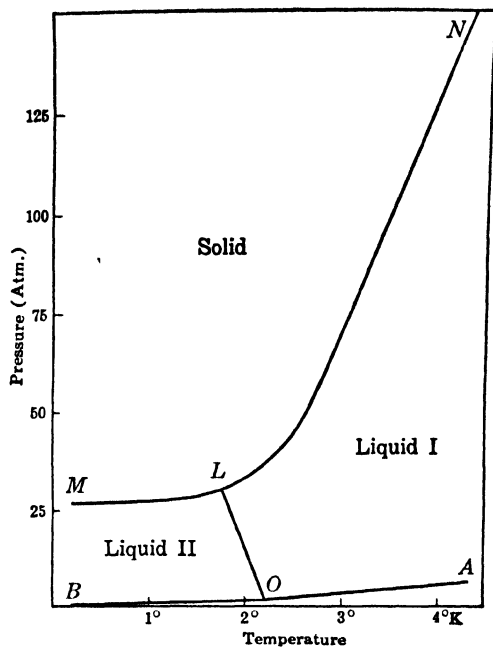


FIG. 112. Phase diagram of helium

tive to the others, are the vapor pressure curves of liquid helium I and II, respectively. At the triple point *O*, which is the λ -point of liquid helium under its own vapor pressure (2.19° K.), the two liquids and vapor are in equilibrium. The variation of the λ -point with pressure is shown by *OL*, and at *L* (1.75° K., 30 atm.) solid begins to form; this is the triple point where solid can coexist with the two liquids. The curve *MLN* shows the influence of pressure on the melting point, and hence indicates the conditions of equilibrium between solid helium and either of the two liquids. It will be noted that *LM* flattens out at low temperatures, so that the melting point is almost independent of pressure. Since there is a small but finite difference of density between solid and liquid, the entropy of fusion, L_f/T , must approach zero (cf. p. 461) as the temperature is lowered. The entropy of liquid II is thus practically the same as that of the solid; this suggests that there is virtually the same degree of order in the liquid II and solid states, and so the two phases have structures which differ little from each other. This fact is important in connection with the theory of liquid structure, for it implies that at low temperatures, when the thermal

movement of the molecules is very small, there is little distinction between liquid and solid.

A curious property of the helium system arises from the fact that liquid cannot exist above the critical temperature (5.2°K.); at pressures above about 150 atm. the melting point exceeds this temperature, and so at high pressures solid helium "melts" to form the gas.²²

LIQUID CRYSTALS

The Mesomorphic State.—When solid cholesteryl benzoate is heated it melts sharply to form a turbid liquid and at a higher temperature an equally sharp change to a clear liquid occurs (F. Reinitzer, 1888). On cooling the changes take place in the reverse direction at the same temperatures. These phenomena were studied by O. Lehmann (1889), who discovered the fact that the turbid liquid is doubly refracting and gives interference patterns in polarized light; since these are properties generally associated with anisotropic crystals (p. 346), Lehmann proposed the name **liquid crystals** for the liquids exhibiting these unusual optical effects. This title is not satisfactory, as the substances have no crystalline properties, for apart from double refraction and related phenomena, they are true liquids possessing surface tension and ability to flow. Various other terms, e.g., crystalline liquids, anisotropic liquids, and paracrystals, have been proposed, but the best (G. Friedel, 1922) is probably **mesomorphic state** (Greek: *intermediate form*). In spite of its disadvantages, however, the original name "liquid crystals" is still frequently employed. It was thought at one time that the turbid liquid contained minute particles in suspension, but these cannot be detected microscopically, neither can they be filtered off, and the mesomorphic liquid appears to be quite homogeneous. The temperature at which the solid melts to form this liquid is usually called the transition point, and the higher temperature, at which the true isotropic liquid is obtained, is referred to as the melting point. The pressure-temperature phase equilibrium diagram is, therefore, similar to that of an enantiotropic substance, e.g., as in Fig. 101; both transition and melting points are affected by pressure (cf. Fig. 103), the extent and direction depending, as for other transitions, on the specific volumes of the two phases in equilibrium.

In addition to cholesteryl benzoate, Lehmann showed that *p*-azoxyanisole and -phenetole, ethyl-*p*-azoxy-benzoate and -cinnamate, and potassium and ammonium oleates form liquid crystals. Subsequently many others were prepared, particularly by D. Vorländer (1908), who obtained nearly 250 mesomorphic compounds. They are all organic substances having long chains, generally terminating in an "active" group, e.g., $-\text{OR}$, $-\text{COOR}$, and often with a mildly active group, e.g., $-\text{C}=\text{C}-$, $-\text{C}=\text{N}-$, $-\text{N}=\text{NO}-$, in the middle. If the compound contains a benzene ring, then only the *para*-derivative, but not the *ortho*- or *meta*-, displays mesomorphism. From a study of these substances it has been possible to distinguish two main classes, called **smectic** (Greek: *soap-like*)

and **nematic** (Greek: *thread-like*) types (G. Friedel, 1922); some substances exhibit both forms, and others show various modifications.

Smectic Phases.—In the smectic state normal liquid flow does not occur; the movement is of a gliding nature, in one plane. On clean glass surfaces there is a tendency to form "stepped drops," and the edge of the liquid often shows a series of fine lines, visible in ordinary light but more clearly seen with polarized light, due to "Grandjean planes." If the drops of liquid are touched the gliding movement is observed. The results suggest a structure consisting of a series of planes, one on top of the other; this is supported by the fact that smectic liquids give X-ray diffraction patterns, but in one direction only. The spacing in this direction, perpendicular to the parallel planes, is little more than the length of a molecule (G. Friedel, 1925; K. Hermann, 1932); for example, for ethyl *p*-azoxybenzoate it is 19.9 Å. in the mesomorphic state compared with 16.2 Å. in the crystal, in which the molecule is probably tilted. When examined in polarized light smectic substances sometimes have a fan-like appearance (focal conics); they are always uniaxial and are not affected by a magnetic field. Some examples of compounds yielding smectic mesophases are quoted in Table 78.

TABLE 78. MESOMORPHIC SUBSTANCES OF THE SMECTIC TYPE

Substance	Transition Point	Melting Point
Ethyl <i>p</i> -azoxybenzoate	114° c.	121° c.
Ethyl <i>p</i> -azoxycinnamate	140°	249°
<i>n</i> -Octyl <i>p</i> -azoxycinnamate	94°	175°
Ammonium Oleate	—	—

Nematic Phases.—Nematic liquids are closer to true anisotropic liquids than are the smectic phases; the former have a low viscosity and flow readily. Mobile thread-like structures can often be seen in thick specimens, especially between crossed Nicol prisms, i.e., in polarized light. The nematic phases are optically uniaxial, and a strong magnetic field is able to orient the axis. If viewed along the lines of force of a magnetic field, the turbid liquid appears clear. Like true liquids, nematic liquids give only diffuse haloes with X-rays, so that the layer structure of smectic phases is not present. The substances mentioned in Table 79 show nematic behavior.

TABLE 79. MESOMORPHIC SUBSTANCES OF THE NEMATIC TYPE

Substance	Transition Point	Melting Point
<i>p</i> -Azoxyanisole	116° c.	135° c.
<i>p</i> -Azoxyphenetole	137°	167°
Anisaldazine	165°	180°
<i>p</i> -Methoxycinnamic acid	170°	186°
Dibenzalbenzidine	234°	260°

Cholesteric Phases.—A number of cholesteryl esters, and a few other optically active substances, show many of the characteristics of nematic behavior, but display in addition special optical properties; these substances are said to form **cholesteric phases**. They show marked color effects in polarized light as a consequence of a layer structure; it differs from that in smectic phases in being very much thicker, viz., 500 to 5000 molecules. The first substance in which mesomorphism was detected, i.e., cholesteryl benzoate, is of the cholesteric type; its transition point is 146° c. and melting point 178.5° c. The *l*-cholesteryl esters from formate to myristate give cholesteric phases, but if the length of the acid chain is increased further, the function of the acidic portion of the molecule predominates and the compounds are smectic.

Compounds with Several Mesophases.—Cholesteryl esters from the caprylate to the myristate, inclusive, give a smectic liquid on melting and this changes sharply, at a higher temperature, to a cholesteric phase which finally gives the clear liquid; for the myristate, for example, the three temperatures are 72°, 78° and 83° c. A few substances are known which exist in smectic and nematic forms, with a definite transition temperature, and some may have more than one smectic phase. Ethyl anisal-*p*-aminocinnamate, for example, undergoes the following changes, starting with the metastable solid.

Crystal \rightleftharpoons Smectic Phase I \rightleftharpoons Smectic Phase II \rightleftharpoons Nematic Phase \rightleftharpoons Liquid
83° 91° 118° 139°

The situation here is complicated by the fact that the stable modification of the solid melts at 108° and so forms smectic phase II directly.

Theory of Liquid Crystals.—The generally accepted views on the structure of liquid crystals are based on the "swarm theory" of E. Bose (1909). Since mesomorphic behavior is restricted to substances having long chains, the molecules will probably tend to arrange themselves parallel to each other. In this way a number of groups or "swarms" would be obtained, in each of which there is a more or less definite orientation, but the arrangement in one swarm will not necessarily be parallel to that in another. In a sense, therefore, the mesomorphic liquid may be regarded as similar to a mass of small crystals; in each crystal there is definite orientation, but the distribution of the crystals is random. The swarms may be compared to the small crystals, although the molecules in the former are not to be regarded as arranged in a rigid manner, neither can the size of the swarm be assumed constant.* The turbidity of liquid crystals is attributed to the scattering of light by the swarms; the effect is similar to that observed when a clear crystalline substance, or glass, is finely powdered. As the temperature is raised the swarms diminish in

* It should be noted that there is a similarity between the swarm theory of liquid crystals and the theory of cybotactic groups in liquids (p. 511); the swarms in liquid crystals would, however, need to be much larger than the groups in a normal liquid.

size, because of increased thermal movements of the molecules, and the melting point, when the liquid clears, is the temperature at which the swarms become too small to scatter light. The double refraction, resulting from orientation within the swarms, will disappear at the same time.

The optical properties of the smectic state suggest that the molecules are arranged in approximately parallel layers; the length of the molecules is roughly perpendicular to these planes, but within each layer the separation between the molecules is not uniform. The arrangement may be compared to a number of hair-brushes, in which the bristles are parallel but set at random, placed one on top of another. In the nematic state, the alignment of the molecules is believed to remain, but the layers no longer exist; there is now freedom of movement in the direction of the length of the molecules. The molecular groups in the nematic phase have been compared to a loose bundle of cigars, so that individuals can move up and down while remaining parallel. The orientations of long molecules in the manner suggested would clearly lead to a difference in optical properties in different directions, and the appearance of double refraction can thus be understood. It is probably significant that substances giving smectic phases have either only one polar group, e.g., ammonium oleate, or a fairly strongly active group at each end, as in the *p*-azoxy-esters; for such compounds layer formation might be expected, and is often found in the solid crystal. When the molecule has a weakly active group both at the end and the middle, as in the *p*-azoxy-ethers, an imbricated structure has been found in the solid state (J. D. Bernal and Miss D. Crowfoot, 1933), the end of one molecule coming opposite the middle of the next. The formation of nematic phases by such substances is not surprising. The structure of cholesteric phases is not yet understood, but it is no doubt connected with the broad, almost flat, sterol residue present in the molecules.

The remarks made above can only be regarded as an outline of the theory of the mesomorphic state, as the phenomena are not always simple. It is not clearly understood why certain long chain molecules which might be expected to form liquid crystals do not do so, although it is important to remember that in certain cases the presence of substituents in the molecule makes the melting point so high that it is well above the hypothetical transition point between crystalline and mesomorphic phases. The potentiality of forming liquid crystals may exist, but it cannot be realized.²³

THE VITREOUS STATE

The Nature of Glass.—When certain liquids are cooled fairly rapidly there is no formation of crystals at a definite temperature, such as occurs on slow cooling. The viscosity of the liquid increases steadily, and finally a glassy substance is obtained. This property of forming a glass when cooled from the molten state is possessed particularly by silica and boron trioxide, and by mixtures of one or both of these substances with oxides

of the alkali or alkaline-earth metals or lead. Other inorganic compounds capable of forming glasses are germanium dioxide, GeO_2 , which is similar to silica, arsenic and phosphoric oxides, arsenious oxide and beryllium fluoride. Among simple organic substances obtainable in the vitreous state, mention may be made of glycerol and glucose. The chief characteristics of a glass are hardness, rigidity and ability to withstand shearing stresses, which are all properties of the solid state. On the other hand, glasses, if not in a state of strain, are optically isotropic, like liquids, and transparent; the apparent solidity is due to high viscosity, for on heating the latter decreases and the glass passes, without any sharp transition, into a mobile liquid. On long standing, especially at a fairly high temperature, glasses undergo "devitrification," when crystals separate just as they do from a supercooled liquid or a supersaturated solution. For the reasons given, some writers regard glasses as amorphous solids, for they possess no definite external structure, but others prefer to consider them as supercooled liquids. The difference between the two points of view is somewhat academic, and the main fact to remember is that the glassy or vitreous state is a condition, in which certain substances can exist, lying between the solid and liquid states.

The Structure of Glass.—Important progress has been made in elucidating the structure of the glassy state, as a result of the theoretical views of V. Goldschmidt (1926) and W. H. Zachariasen (1932), and of the X-ray diffraction experiments made by B. E. Warren, *et al.*, (1933, *et seq.*). It had been known for some time that glasses give diffuse diffraction bands, like liquids, and J. T. Randall (1930) noted that the position of the band obtained from vitreous silica indicated a spacing of diffraction centers (4.33 Å) close to that in the crystalline form of silica, α -cristobalite (4.11 Å). A similar agreement was obtained in the comparison of other substances in the glassy and crystalline states, and it was suggested that the glass was made up of minute crystalline particles, or "crystallites," of α -cristobalite with linear dimensions of about 15 Å. Apart from the small difference in the spacings in the two forms of silica, the theory meets with several difficulties. More recent work on the intensity of the X-ray scattering has shown that the crystallites, if they exist, are probably not more than about 8 Å. in size, which is little larger than a unit cell of cristobalite; this can hardly be regarded as a separate crystallite. The diffuse X-ray halo of vitreous silica passes over sharply to the narrow band of α -cristobalite on heat treatment; if small crystals were initially present, they would grow in size gradually and the transition would not be sudden. Further, cristobalite undergoes a volume change at 200° to 300° C., but vitreous silica does not. Finally, a mass of crystallites arranged at random would give appreciable X-ray scattering at small angles, like a gas, but actually none is observed. Dried silica gel, which consists of an aggregate of small crystals, does give small angle scattering.

By applying the radial distribution method, described on p. 505, to vitreous silica and boron trioxide, B. E. Warren has shown that the structural unit, viz., SiO_4 , with four oxygen atoms arranged tetrahedrally round the silicon, or BO_3 , with three oxygen atoms in an equilateral triangle about the boron, is almost identical with that in the corresponding crystalline substance. The difference between the crystal and the glass is that in the former these groups, i.e., SiO_4 or BO_3 , are arranged regularly and uniformly throughout the whole of the solid, but in the glass the distribution is random. This agrees with the suggestions made by Zachariasen, whose views are represented by a two-dimensional diagram in Fig. 113; (A) represents an oxide of composition X_2O_3 in the crystalline form, and (B) is a projection of the same oxide in the vitreous state. The units of structure are the same; an observer on the atom X (black circles), only able to see as far as the adjoining atoms in each direction, could not distinguish between the crystal and the glass. Each X atom is surrounded by three oxygens, and each oxygen atom is joined to two X atoms, in both crystalline and vitreous states. It will be seen, however, that in the crystal the two X atoms coordinated to the oxygen fall on a straight line passing through its center, but this is not true of the glass. The result is that in the latter rings of different sizes can form, and the identical repetition of structure, which is the fundamental property of the crystal, is lost. The essential similarity between crystalline and vitreous states means that the potential energies will not differ greatly, and so the relative stability of the latter can be understood. The structures shown in Fig. 113 may be taken to be those of boric oxide. In silica the same type of arrangement would exist, except that each silicon atom would be associated tetrahedrally with four oxygen atoms, each of which is attached to two silicon atoms, so that the structure would extend in three dimensions, instead of two as shown.

In soda-silica glasses the sodium ions fit into the holes in the silicon-oxygen network; the radial distribution results show that each sodium is, on the average, surrounded by six oxygen atoms at a distance, between centers, of 2.35 Å. (B. E. Warren, 1938). The addition of sodium oxide to the silica results in a partial breakdown of the continuous system of bonds present in silica itself; although the main structure is unchanged, some of the oxygen atoms are joined to only one silicon atom, instead of two. This is a direct consequence of the introduction of the basic oxide, as may be seen by comparing the structure of crystalline silica (p. 390) with that of a silicate ion (p. 395). The incompleteness of some of the silicon-oxygen rings accounts for the softening temperature of a soda-silica glass being less than that of vitreous silica. It is important to remember, however, that the essential structure of the latter must persist to a great extent, otherwise the glassy character would be lost.

For a substance, e.g., an oxide, to form a glass, or to be the basis for the production of a glass, it should be able to exist in a two- or three-dimensional network lacking in periodicity, but nevertheless having an

energy content little different from that of the crystal. In general, it may be said that the units of structure should be the same in the vitreous and crystalline states, but their arrangement will be regular in the former and irregular in the latter. Zachariasen has shown that three simple

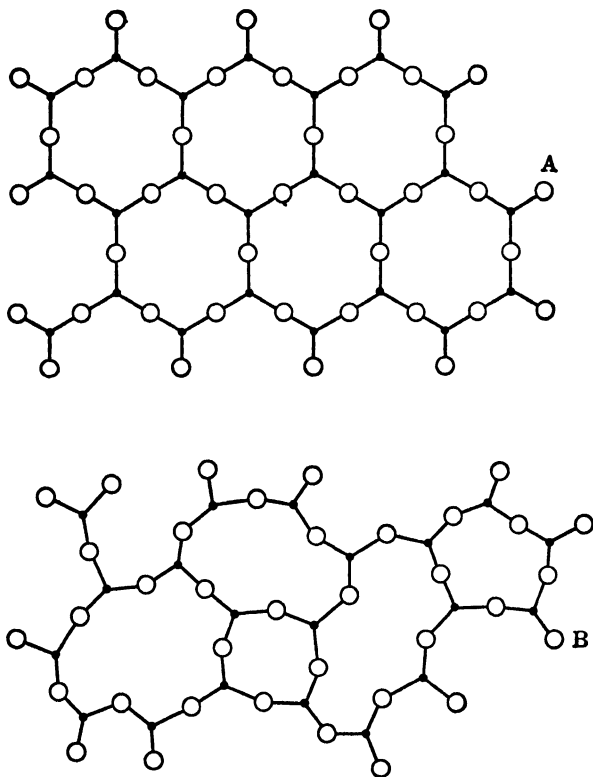


FIG. 113. Oxide X_2O_3 ; crystalline form (A) and vitreous form (B)

rules must be obeyed if these conditions are to hold; they are (i) an oxygen atom is linked to not more than two X atoms; (ii) the number of oxygen atoms surrounding each X atom must be small, e.g., 3 or 4; and (iii) the groups of oxygen atoms, i.e., triangles or tetrahedra, share corners only, not edges or faces. If the network is to extend in three dimensions, at least three corners of the groups of oxygen atoms must be shared with other groups. All the oxides mentioned on p. 519 as forming glasses satisfy these conditions; so also does beryllium fluoride, for in the crystal every beryllium atom is surrounded by four fluorines, each of which is shared by two beryllium atoms, and this substance does in fact form a glass.

The Formation of a Glass.—Since silicon is quadrivalent each atom should be associated with two oxygen atoms; the fact that in the crystal each silicon is surrounded by four oxygens, makes it inevitable that every oxygen atom should be shared by two silicon atoms. The same conclusion holds for boron trioxide, in which the crystal coordination number of the boron is three. If there is any tendency for the crystal structure to exist in the liquid, as there certainly will be near the melting point, it is clear that complex aggregates will form wherein each oxygen atom holds together two silicon, or two boron, atoms. In these complexes the atoms are bound relatively tightly, and on rapid cooling there is no time for the attachments to be broken so that the groups can arrange themselves in the regular manner required for the crystalline form. As the temperature falls, and the viscosity of the system increases, the possibility of rearrangement decreases, and eventually the random distribution of the liquid state becomes more or less fixed in the glass. The latter could not be formed, however, if the atomic coordination present in the crystal did not persist to some extent in the liquid.

The simplest organic glass-forming substances contain several hydroxyl groups in the molecule; there is little doubt, from viscosity and other data, that there is considerable complexity in the liquid phase, the molecules being held together by a network of hydrogen bonds. On cooling, there is considerable reluctance for these bonds to break and for the molecules to move to the correct positions required for the formation of a crystal lattice; this reluctance is favored by the high viscosity of the system, and so the whole liquid sets to a glassy mass.²⁴

REFERENCES

- Adam, N. K. "The Physics and Chemistry of Surfaces."
 Bingham, E. C. "Fluidity and Plasticity."
 Hatschek, E. "Viscosity of Liquids."
 Hildebrand, J. H. "Solubility."
 Randall, J. T. "The Diffraction of X-Rays and Electrons, etc."
 Sugden, S. "The Parachor and Valency."
1. Lewis, *Trans. Faraday Soc.*, **7**, 94 (1911); Hildebrand, *J. Am. Chem. Soc.*, **43**, 500 (1921); *op. cit.*
 2. Harkins, *J. Am. Chem. Soc.*, **44**, 2665 (1922); *J. Chem. Phys.*, **8**, 552 (1941); *Chem. Rev.*, **29**, 385 (1941).
 3. Antonoff, *J. Chim. phys.*, **5**, 372 (1907); *Kolloid-Z.*, **59**, 7 (1932); **64**, 336 (1933); *J. Phys. Chem.*, **46**, 497 (1942); **47**, 709 (1943); Yoffe and Heymann, *ibid.*, p. 409; Cupples, *ibid.*, **48**, 75 (1944); Reynolds, *J. Chem. Soc.*, **119**, 466 (1921); Bartell, *et al.*, *J. Am. Chem. Soc.*, **55**, 2769 (1933); Carter and Jones, *Trans. Farad. Soc.*, **30**, 1027 (1934).
 4. Ramsay and Shields, *Z. physik. Chem.*, **12**, 450 (1893); Sugden, *J. Chem. Soc.*, **99**, 1483 (1921); Richards and Carver, *J. Am. Chem. Soc.*, **43**, 827 (1921); Jones and Ray, *ibid.*, **59**, 187 (1937).
 5. Jaeger, *Z. anorg. Chem.*, **101**, 1 (1917); Sugden, *J. Chem. Soc.*, **121**, 858 (1922); **125**, 27 (1924); Long and Nutting, *J. Am. Chem. Soc.*, **64**, 2476 (1942).
 6. Traube, *Ber.*, **20**, 2644 (1887); Harkins and Brown, *J. Am. Chem. Soc.*, **41**, 499 (1919); Freud and Harkins, *J. Phys. Chem.*, **23**, 1217 (1929).

7. Du Noüy, *J. Gen. Physiol.*, **1**, 521 (1918-19); Harkins and Jordan, *J. Am. Chem. Soc.*, **52**, 1751 (1930); Freud and Freud, *ibid.*, p. 1772; Adam and Shute, *Trans. Farad. Soc.*, **34**, 758 (1938); Andreas, Hauser and Tucker, *J. Phys. Chem.*, **42**, 1001 (1938); Bartell and Davis, *ibid.*, **45**, 1321 (1941); Smith, *ibid.*, p. 671; **48**, 168 (1944); Tartar, *et al.*, *J. Am. Chem. Soc.*, **62**, 2375 (1940); Dole and Swartout, *ibid.*, p. 3039; McBain and Sharp, *ibid.*, **63**, 1422 (1941).
8. Bartell, *et al.*, *ibid.*, **50**, 1961 (1928); **54**, 936 (1932); **55**, 2419 (1933); Harkins and Humphrey, *ibid.*, **38**, 236 (1916); Speakman, *J. Chem. Soc.*, 1449 (1933).
9. Eötvös, *Ann. Phys.*, **27**, 448 (1886); Ramsay and Shields, *Phil. Trans. Roy. Soc.*, **184A**, 647 (1893); *J. Chem. Soc.*, **63**, 1089 (1893); Walden and Swinne, *Z. phys. Chem.*, **82**, 271 (1913); Hunte and Maass, *J. Am. Chem. Soc.*, **51**, 153 (1929); Gallagher and Hibbert, *ibid.*, **59**, 2514 (1937).
10. Macleod, *Trans. Farad. Soc.*, **19**, 38 (1923); Ferguson, *ibid.*, p. 408; Sugden, *J. Chem. Soc.*, **125**, 1177 (1924); *op. cit.*; Fowler, *Proc. Roy. Soc.* **159A**, 227 (1937); Lewis, *J. Phys. Chem.*, **44**, 1007 (1940).
11. Traube, *Annalen*, **265**, 27 (1891); Langmuir, *J. Am. Chem. Soc.*, **39**, 1883 (1917).
12. Washburn and Williams, *ibid.*, **35**, 739 (1913); Bingham, *Ind. Eng. Chem.*, **6**, 233 (1914); Bingham, *et al.*, *J. Am. Chem. Soc.*, **38**, 27 (1916); Gemant, *J. App. Phys.*, **13**, 22 (1942).
13. Gibson and Jacobs, *J. Chem. Soc.*, **117**, 473 (1920); Symmes and Lautz, *Ind. Eng. Chem. (Anal.)*, **1**, 35 (1929); Barr, *Phil. Mag.*, **1**, 395 (1926).
14. Batschinski, *Z. phys. Chem.*, **84**, 643 (1913); Bingham and Kenney, *J. App. Phys.*, **11**, 192 (1942).
15. Kraemer and Lansing, *J. Phys. Chem.*, **39**, 153 (1935); *Ind. Eng. Chem.*, **30**, 1200 (1938); Huggins, *J. Phys. Chem.*, **42**, 911 (1938); **43**, 439 (1939); *J. App. Phys.*, **10**, 700 (1939); *J. Am. Chem. Soc.*, **64**, 2716 (1942); Eirich, *Phys. Soc. Rep.*, **7**, 329 (1940); Gemant, *J. App. Phys.*, **13**, 90 (1942); Simha, *ibid.*, p. 147; Lauffer, *Chem. Rev.*, **31**, 561 (1942); Flory, *J. Am. Chem. Soc.*, **65**, 372 (1943); Bartovics and Mark, *ibid.*, p. 1901; Meyer, "Natural and Synthetic High Polymers."
16. Eyring, *J. Chem. Phys.* **4**, 283 (1936); Ewell and Eyring, *ibid.*, **5**, 726 (1937); Ewell, *ibid.*, p. 571; Kauzmann and Eyring, *J. Am. Chem. Soc.*, **62**, 3113 (1940); Roseveare, Powell and Eyring, *J. App. Phys.*, **12**, 669 (1941); Gemant, *ibid.*, p. 827; Fürth, *Proc. Camb. Phil. Soc.*, **37**, 252, 276, 281 (1941).
17. Smith, *Z. phys. Chem.*, **57**, 685 (1907); Hammick, *J. Chem. Soc.*, 273 (1930); Blatchford, *Proc. Phys. Soc.*, **45**, 493 (1933); Gingrich, *J. Chem. Phys.*, **8**, 29 (1940); Bacon and Fanelli, *J. Am. Chem. Soc.*, **65**, 639 (1943); Powell and Eyring, *ibid.*, p. 648.
18. Stewart, *Phys. Rev.*, **30**, 232 (1927); Zernike and Prins, *Z. Phys.*, **41**, 184 (1927); Debye and Menke, *Phys. Z.*, **31**, 797 (1930); **33**, 593 (1932); for reviews, see Randall, *Ann. Rep. Chem. Soc.*, **34**, 169 (1937); *op. cit.*; Gingrich, *Rev. Mod. Phys.*, **15**, 90 (1943).
19. Prins, *Z. Phys.*, **56**, 617 (1929); Stewart, *Phys. Rev.*, **31**, 174 (1928); **32**, 153 (1929); **33**, 889 (1929); Bernal and Fowler, *J. Chem. Phys.*, **1**, 515 (1933); Ward, *ibid.*, **2**, 153 (1934); Katzoff, *ibid.*, p. 841; Warren, *et al.*, *Phys. Rev.*, **44**, 969 (1933); *J. Chem. Phys.*, **4**, 236 (1936); **6**, 666 (1938); *J. App. Phys.*, **8**, 645 (1937); Harvey, *J. Chem. Phys.*, **6**, 111 (1938); **7**, 878 (1939); Zachariasen, *ibid.*, **3**, 158 (1935); Gingrich, *et al.*, *ibid.*, **8**, 29 (1940); **10**, 504 (1942); **11**, 351 (1943); *Rev. Mod. Phys.*, **15**, 90 (1943); Gamertsfelder, *J. Chem. Phys.*, **9**, 450 (1941); Campbell and Hildebrand, *ibid.*, **11**, 330, 334 (1943); Pierce, *ibid.*, **3**, 252 (1935); **5**, 717 (1937); *J. Am. Chem. Soc.*, **60**, 779 (1938).
20. Stewart, *Chem. Rev.*, **6**, 483 (1929); *Rev. Mod. Phys.* **2**, 116 (1930); *Trans. Farad. Soc.*, **29**, 982 (1933); Bernal, *ibid.*, **33**, 27 (1937); Prins, *ibid.*, **110**; see also, Hildebrand, *Science*, **90**, 1 (1939); *Proc. Phys. Soc.*, **56**, 221 (1944).
21. For reviews, see Lennard-Jones, *Proc. Phys. Soc.*, **52**, 729 (1940); Glasstone, "Theoretical Chemistry," Chap. IX.

22. For reviews, see Burton, *J. App. Phys.*, **9**, 489 (1938); Allen and Jones, *Nature*, **143**, 227 (1939); Darrow, *Rev. Mod. Phys.*, **12**, 257 (1940).
23. Reinitzer, *Monatsh.*, **9**, 435 (1888); Lehmann, *Z. phys. Chem.*, **4**, 468 (1889); **102**, 91 (1922); *Ann. Phys.*, **2**, 649 (1900); Bose, *Phys. Z.*, **10**, 32, 230 (1909); Friedel, *Ann. Physique*, **18**, 273 (1922); Kast, *Phys. Z.*, **38**, 627 (1937); Fürth and Sitte, *Ann. Phys.*, **30**, 388 (1937); Kast, *Z. Elek.*, **45**, 184 (1939); for general discussions, see *Z. Krist.*, **79**, 1 *et seq.* (1931) (cf. *Ann. Rep. Chem. Soc.*, **28**, 280 (1931)); *Trans. Farad. Soc.*, **29**, 881 *et seq.* (1933).
22. Zachariasen, *J. Am. Chem. Soc.*, **54**, 3841 (1932); *Phys. Rev.*, **39**, 185 (1932); *J. Phys.*, **3**, 162 (1935); Hägg, *ibid.*, p. 42; Warren, *et al.*, *Z. Krist.*, **86**, 349 (1933); **89**, 481 (1934); *Phys. Rev.*, **45**, 657 (1934); *J. Am. Ceram. Soc.*, **17**, 249 (1934); **18**, 269 (1935); **19**, 202 (1936), **21**, 49, 259 (1938); *Chem. Rev.*, **26**, 237 (1940); *J. App. Phys.*, **13**, 602 (1942); Huggins, *J. Phys. Chem.*, **47**, 502 (1943).

CHAPTER VIII

PHYSICAL PROPERTIES AND MOLECULAR STRUCTURE

ADDITIVE AND CONSTITUTIVE PROPERTIES

Physical properties of a system, compound or mixture, may be divided roughly into three categories. An **additive** property is one which, for a given system, is the sum of the corresponding properties of the constituents. The only strictly additive property is mass, for the mass of a molecule is exactly equal to the sum of the masses of its atoms, and similarly the mass of a mixture is the sum of the separate masses of the constituent parts. There are other properties, however, which are approximately additive, e.g., the molar volume, as will be seen shortly. A property is said to be **constitutive** when it depends primarily on the arrangement of the atoms in the molecule, and to a lesser extent on their nature and number. Many physical properties are partly additive and partly constitutive; in some, e.g., refraction, the former is predominant, whereas in others, e.g., boiling point, the latter factor is more important. A **colligative** property is one depending primarily on the number of molecules concerned and not on their nature; the volume of a gas, at constant temperature and pressure, is a simple illustration, but it is in the study of solutions that these properties are chiefly encountered. In the present chapter some additive and constitutive properties are described, whereas colligative, sometimes called osmotic, properties of solutions will be considered in Chapter IX.

MOLAR VOLUME

Additivity of Molar Volume.—The first physical property showing marked additivity to be studied was the molecular, or molar, volume, i.e., the volume occupied by a mole of liquid; it is obtained by multiplying the molecular weight by the specific volume, or by dividing it by the density, of the liquid. H. Kopp (1842, 1855) noted that molar volumes of organic compounds, measured at their respective boiling points* under atmospheric pressure, were to a great extent additive functions of the volume equivalents of their constituent elements. Although Kopp's choice of boiling points as the temperatures for the comparison of molar volumes was empirical, it is now known to have a theoretical basis, for boiling points are corresponding temperatures (p. 440). Since isomeric compounds of similar constitution, e.g., methyl propionate, ethyl acetate and propyl formate, have almost identical molar volumes at their

* Kopp did not actually determine densities at the boiling points, but at lower temperatures and extrapolated to the boiling point.

boiling points, and since in various homologous series there is a constant difference of 22 cc. for each CH_2 group, it appears that the molecular volume is essentially an additive property. It was evident in Kopp's later work, and this has been confirmed, that constitutive factors are also involved; thus, the oxygen atom contributes different amounts to the molar volume according as it is part of a hydroxyl or a carbonyl group.

The volume equivalents of various elements may be estimated in the following manner: assuming the CH_2 group to contribute 22.0 cc., which is the average found in a number of homologous series, the volume equivalent of 2H is estimated by subtracting $n \times 22$ for n CH_2 groups from the molar volume of a paraffin $\text{C}_n\text{H}_{2n+2}$; the mean value is 11.0, so that each hydrogen atom contributes 5.5 cc. and hence 11.0 cc. must be the volume equivalent of a carbon atom. With these values known it is a relatively simple matter to determine those for other elements, and some of the results obtained by Kopp are given in Table 80; the symbol

TABLE 80. VOLUME EQUIVALENTS OF ELEMENTS

Atom	Kopp	Le Bas	Atom	Kopp	Le Bas
Hydrogen	5.5	3.7	Iodine	37.5	37.0
Carbon	11.0	14.8	Oxygen ($-\text{O}-$)	7.8	7.4
Chlorine	22.8	22.1	Oxygen ($\text{O}=\text{}$)	12.2	12.0
Bromine	27.8	27.0	Sulfur	22.6	

$-\text{O}-$ refers to the oxygen atom in hydroxyl groups and in ethers, and $\text{O}=\text{}$ in the carbonyl group. Somewhat different equivalents were calculated by W. C. Lossen (1889) and by G. Le Bas (1912); those given by the latter, also included in the table, require 15.0 cc. to be subtracted for a benzene ring. To illustrate the application of the equivalents in Table 80 to test the validity of the additivity of molar volumes at the boiling point, the values for ethyl benzoate may be calculated and compared with that observed (174.6 cc.). In spite of the good agreement

	Kopp		Le Bas
9C	99.0	9C	133.2
10H	55.0	10H	37.0
$-\text{O}-$	7.8	$-\text{O}-$	7.4
$\text{O}=\text{}$	12.2	$\text{O}=\text{}$	12.0
		C_6H_5 ring	-15.0
			<hr/>
	174.0 cc.		174.6 cc.

in this and other instances, there are many cases where structural factors have some influence, and although the figures in Table 80 are useful for the purpose of calculating approximate molecular volumes, it is unsafe to use the measured molar volume at the boiling point to throw light on the constitution of the molecule. As a general rule associated liquids, e.g., alcohols, do not show any unusual abnormality of behavior; this may be ascribed to the diminished extent of association at the boiling point, or to the small contribution of the hydrogen bond (see p. 509).

THE PARACHOR

Surface Tension and Density.—It has been seen (p. 494) that the surface tension γ of a liquid is related to its density ρ and that of the vapor ρ' at the same temperature by Macleod's equation

$$\frac{\gamma^{\frac{1}{3}}}{\rho - \rho'} = C, \quad (1)$$

which holds over a wide range of temperature. If both sides are multiplied by the molecular weight it follows that

$$\frac{M\gamma^{\frac{1}{3}}}{\rho - \rho'} = \text{a constant}, [P]. \quad (2)$$

The resulting constant $[P]$ has been called the **parachor** (S. Sugden, 1924), a name which implies that it is a comparative or relative volume. If the density of the vapor ρ' is neglected in comparison with that of the liquid, it follows that

$$[P] = M\gamma^{\frac{1}{3}}/\rho = V_m\gamma^{\frac{1}{3}}, \quad (3)$$

where V_m is the molar volume. If the temperature is such that the surface tension is unity, then $[P]$ is equal to V_m , so that the parachor may be regarded as the molar volume of a substance when its surface tension is unity. A comparison of the parachors of different substances is thus equivalent to the comparison of molar volumes under such conditions that the liquids have the same surface tension; since the molecular attractions, and hence the internal pressures, are then approximately equal it is to be expected that the parachor should provide a better basis for comparing molar volumes than measurement at the boiling points.

From an examination of the data for isomeric substances and of members of homologous series, S. Sugden (1924) showed the parachor to be primarily an additive, and also a constitutive, property. By the procedure described for molar volumes, parachor equivalents have been determined for a number of elements and also for various structural factors; some of the results are recorded in Table 81. It was considered

TABLE 81. PARACHOR EQUIVALENTS *

Carbon	4.8	Sulfur	48.5	Nitrogen	12.5
Hydrogen	17.1	Chlorine	53.8	Double bond	23.2
Oxygen	20.0	Bromine	68.0	Triple bond	46.6
O ₂ in esters	60.0	Iodine	90.0	6-membered ring	6.1

* Several attempts have been made to revise the parachor equivalents, particularly those of carbon and hydrogen; values from 4.8 to 11.5 have been proposed for C, and from 17.1 to 14.4 for H. The fact of the matter is that the parachor is not strictly additive, and constitutive allowance should be included. The contribution of the CH₂ group, for example, varies with the nature of the compound, and with the position of the group in the molecule. The equivalents in Table 81 are thus to be regarded as no more than a moderately satisfactory compromise for purposes of simplicity. For compounds of low molecular weight the values C 7.2 and H 16.2 may be preferred.

at one time that the presence of a semipolar bond (p. 99) was accompanied by a decrease of 1.6 units in the parachor, but some doubt is now felt as to whether this correction is necessary.

The additivity of the equivalents in Table 81 may be tested by calculating the parachor for toluonitrile, $\text{CH}_3\text{C}_6\text{H}_4\text{CN}$; this contains eight carbon atoms (8×4.8), seven hydrogen atoms (7×17.1), and one nitrogen atom (12.5), and in addition there are the following structural factors, one triple bond in the nitrile group (46.6), three double bonds in the benzene ring (3×23.2) and the six-membered ring (6.1). The sum of these amounts is 292.9, which is in excellent agreement with the experimental values, viz., *ortho*- 290.6, *meta*- 295.6, *para*- 294.4.

The parachor equivalents recorded apply to the atoms in their normal valence states; any increase of valence is generally accompanied by a decrease of approximately 13 in the parachor for each unit of valence. The parachor defect is somewhat less for elements of high atomic number, e.g., 11 for selenium and tellurium, and 8.6 for iodine. The observed parachor of sulfur hexafluoride is 143.3; this is made up of 150 for six fluorine atoms and 48.5 for the bivalent sulfur atom, from which must be deducted 4×13 for the increase of valence from two to six, making a total of 146.5.

Associated liquids, e.g., alcohols and carboxylic acids, have parachors which generally are not constant, but increase somewhat with temperature; the values are a few units lower than those to be expected from additivity considerations. For example, at 20° c. the parachor of ethyl alcohol is 126.9 and it increases to 131.0 at 200° c., compared with the additive value of 132.2. The small discrepancy may be attributed to the presence of hydrogen bonds; since the hydrogen is then virtually bivalent it may possibly be accompanied by a decrease in the parachor. The extent of association decreases with increasing temperature and so the parachor approaches the additive value. Other compounds containing hydrogen bonds, e.g., β -diketones and derivatives of *o*-phenols, also exhibit parachor defects.

Applications.—In view of the appreciable contributions made by structural factors, it is evident that the parachor might be used in certain cases to decide between alternative configurations; one or two examples may be considered. Acetaldehyde forms a liquid polymer, paraldehyde, having the formula $(\text{C}_2\text{H}_4\text{O})_3$; its parachor is 298.7, whereas three molecules of acetaldehyde would correspond to 3×121.2 , i.e., 363.6. The decrease of $363.6 - 298.7$, i.e., 64.9, is approximately equivalent to the parachor contribution of three double bonds; it appears, therefore, that in polymerization the double bonds of the acetaldehyde have been eliminated. The generally accepted formula for paraldehyde, involving a six-membered ring containing three oxygen and three carbon atoms, with no double bonds, has a calculated parachor of 300.1, which is close to the observed value.

Much controversy has centered about the structure of the isocyanide radical, $-\text{NC}$, and important light has been thrown on the subject by parachor measurements. The mean parachor equivalent of the $-\text{NC}$ group, in a number of alkyl and aryl isocyanides, is 66; this corresponds

closely to that required for the structure $-N \equiv C^-$, thus N(12.5), C(4.8), triple bond (46.6), making a total of 63.9. The alternative structure $-N=C$ would have a parachor equivalent of only 40.6. The modern view is that in the isocyanide group resonance (p. 111) occurs between the two structures, but the configuration involving the triple bond predominates. This conclusion is supported by evidence from spectra, and dipole moments. The structure of carbon monoxide presents an analogous problem; the observed parachor 61.6 is in harmony with the formula $-C \equiv O^+$, which should have a value of 71.4, whereas the parachor for $C=O$ would be 48.0. Both configurations probably contribute to the resonance, the one with the triple bond being the more important.

Parachor Anomalies.—It must be pointed out that the use of the parachor for the solution of structural problems is not free from objections. For example, parachor measurements indicated that organic azides, containing the $-N_3$ group, and aliphatic diazo-compounds have cyclic structures. It appears to be proved by electron diffraction measurements (p. 595), however, that the groups are actually linear. The erroneous conclusions may perhaps have been due to the failure to take into account the possibility of resonance. Further, serious parachor anomalies exist in several organometallic compounds; in the dialkyl sulfide derivatives of palladous chloride, having the general formula $(R_2S)_2PdCl_2$, the parachor equivalent of the palladium atom decreases from 36, when R is methyl, to -7 when R is amyl. A similar steady decrease in the contribution of the metal atom occurs in the mercury mercaptides, $Hg(SR)_2$, and in other series of compounds (F. G. Mann and D. Purdie, 1935-6), as the length of the hydrocarbon chain is increased. It is probable that as the central atom becomes more completely covered with alkyl groups, it becomes less and less capable of making its contribution to forces of cohesion and hence to surface tension. In such circumstances the normal additivity relationships of the parachor might be expected to fail.¹

MOLAR REFRACTION

Refractive Index and Density.—In recent years the molar refraction has acquired an important significance which will be considered more fully later; in the meantime reference may be made to the additive and constitutive nature of this property. In 1858, J. H. Gladstone and T. P. Dale found empirically that the quantity $(n - 1)/\rho$, where n is the refractive index and ρ the density of a liquid, was almost independent of temperature for light of a definite wave length. The product of this quantity and the molecular weight was called the **molar refraction**, and this proved to be a property which was partly additive and partly constitutive (J. H. Gladstone and T. P. Dale, 1863; H. Landolt, 1864). The deduction of a theoretical relationship between refractive index and density, which should be constant at all temperatures, viz.,

$$\frac{n^2 - 1}{n^2 + 2} \cdot \frac{1}{\rho} = \text{constant}, (r), \quad (4)$$

where r is the **specific refraction**, was made independently by H. A. Lorentz (1880) and L. V. Lorenz (1880), from the electromagnetic and wave theories of light, respectively; this has rendered obsolete the empirical equation of Gladstone and Dale. The molar refraction of a compound is now defined by the expression

$$[R] = \frac{n^2 - 1}{n^2 + 2} \cdot \frac{M}{\rho}, \quad (5)$$

the value depending only on the wave length of the light used for measuring the refractive index. In addition to having a theoretical basis, it has the advantage over the older formula of being almost the same in both liquid and gaseous states, so that the molar refraction may be regarded as a fundamental property of the compound. Careful observation has shown that the value for the liquid is slightly less than for the gas, but the difference is rarely more than a few per cent. Almost the whole of the work described below has been carried out with liquids, since measurements can then be made more easily. According to theoretical considerations (cf. p. 542), based on the assumption that molecules are perfectly conducting spheres, the molar refraction for radiation of infinite wave length, i.e.,

$$[R]_{\infty} = \frac{n_{\infty}^2 - 1}{n_{\infty}^2 + 2} \cdot \frac{M}{\rho}, \quad (6)$$

represents the *actual* volume of the molecules in 1 mole of substance, as distinct from the apparent volume M/ρ . For purposes of comparison of true molar volumes, therefore, the quantity $[R]_{\infty}$ should be used; although molecules are not conducting spheres they may be assumed to be so as a first approximation. Measurement of refractive index with light of very long wave length, e.g., infrared radiation, is difficult, and extrapolation of observations made with visible light gives incorrect results (p. 536); it is the practice, therefore, to compare molar refractions for light of a definite wave length, e.g., α -, β - or γ -lines of hydrogen, or the sodium- D line. This procedure is obviously far from ideal, but it has proved to be a satisfactory approximation.

Like the other forms of molar volume the molar refraction is an additive and constitutive property. The early calculations in this connection were made by J. W. Bruhl (1880 *et seq.*) and subsequently by F. Eisenlohr (1910), whose results for the refraction equivalents of various atoms and structural units are recorded in Table 82*; the suffixes refer to the spectral lines mentioned above. Refraction equivalents have also been estimated for nitrogen and sulfur, but as the values are apparently different for every form of combination, e.g., F. Eisenlohr (1913) mentions thirty different equivalents for nitrogen, the results do not seem to have any fundamental significance.

* Somewhat different values are given by W. Swietoslawski (1920).

TABLE 82. REFRACTION EQUIVALENTS

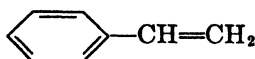
	$[R]_a$	$[R]_D$	$[R]_B$	$[R]_T$
Carbon	2.413	2.418	2.438	2.466
Hydrogen	1.092	1.100	1.115	1.122
Oxygen (in CO group)	2.189	2.211	2.247	2.267
Oxygen (in ethers)	1.639	1.643	1.649	1.662
Oxygen (in OH group)	1.522	1.525	1.531	1.541
Chlorine	5.933	5.967	6.043	6.101
Bromine	8.803	8.865	8.999	9.152
Iodine	13.757	13.900	14.224	14.521
Double bond	1.686	1.733	1.824	1.893
Triple bond	2.328	2.398	2.506	2.538

It should be emphasized that although the refraction equivalents in Table 82 are recorded to three places of decimals, they are nevertheless approximate because molar refractivities, at least those measured at finite wave lengths, are not strictly additive. The average increment of $[R]_D$ for a CH_2 group varies from 4.606 in a homologous series of esters to 4.634 in the alcohol series; Eisenlohr has used a mean value of 4.618 to evaluate the data given above, and so the equivalents for carbon and hydrogen may be as much as 0.016 in error.

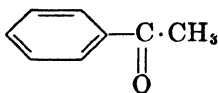
The following results indicate the nature of the agreement between the observed molar refractions for the D -line and those calculated from the data in Table 82.

	cyclo- C_6H_{12}	C_6H_6	$(\text{C}_2\text{H}_5)_2\text{O}$	$(\text{CH}_3)_2\text{CO}$	CHCl_3
Obs.	27.71	26.15	22.48	16.15	21.40
Calc.	27.67	26.31	22.31	16.07	21.42

Anomalous Results.—When an open chain molecule contains a system of conjugated double bonds, the molar refraction is different from that calculated from the known equivalents; this behavior is referred to as **optical anomaly**, and since the observed value is generally the higher, the substance is said to exhibit **optical exaltation**. For example, the observed molar refraction $[R]_D$ of *iso*-diallyl, $\text{CH}_3-\text{CH}=\text{CH}-\text{CH}=\text{CH}-\text{CH}_3$, containing conjugated double bonds, is 1.76 units higher than the calculated value, whereas for the isomeric diallyl, $\text{CH}_2=\text{CH}-\text{CH}_2-\text{CH}_2-\text{CH}=\text{CH}_2$, in which the double linkages are not conjugated, the observed value is smaller by about 0.12 unit. If the conjugated double linkages constitute a closed ring, as in benzene and cyclo-octatetrene, the optical exaltation disappears, but if the conjugated system is partly within the ring and partly in a side chain, as in styrene (I) and acetophenone (II), the positive anomaly appears.



I



II



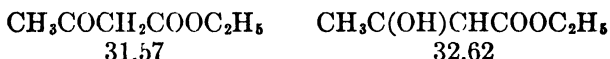
III

Similarly, if there is a gap in the ring, as in α -phellandrene (III), the

optical exaltation is present. Naphthalene and anthracene and other polynuclear systems exhibit marked exaltation, although aromatic ring formation alone has little influence on the molar refractivity. The molar refractions of cyclopropane and cyclobutane and their derivatives are about 0.7 and 0.5 unit, respectively, larger than the additive values.

Applications.—Measurements of refractive index have been used in several cases, particularly among the terpene compounds, to provide some information concerning molecular structure. It has been possible, for instance, to distinguish between the isomeric α - and β -terpenes (K. von Auwers and F. Eisenlohr, 1910). The molecules contain two double bonds, and as the α -form exhibits an optical exaltation of over 1 unit it presumably contains a conjugated system.

Another interesting application of molar refraction is connected with the subject of keto-enolic tautomerism; the two forms should have different values of $[R]$ since the enol-form contains a double bond, and consequently an approximate indication can be obtained whether one or the other predominates in the compound in its normal state. The possibility of optical anomalies, however, renders the conclusions not entirely free from objection. The calculated molar refractions, for the *D*-line, of the keto- and enol-forms of acetoacetic ester are 31.57 and 32.62 respectively; thus,



In the enolic form the double bonds are conjugated and so there should be optical exaltation; comparison of the observed and calculated values of $[R]_D$ for β -ethoxycrotonic ester indicates that this should be about 1.8, so that the molar refraction of the enolic form should be 34.42. The measured molar refraction of acetoacetic ester in its normal state is 32.00, and so the proportion of enolic form should be given by

$$\frac{32.00 - 31.57}{34.42 - 31.57} \times 100 = 15 \text{ per cent.}$$

The result is somewhat higher than that obtained by other methods, and it has been suggested in explanation that the calculated $[R]_D$ for the keto-form is about 0.2 unit too high. It must be remembered, too, that in the enolic form ethyl acetoacetate has a ring structure involving a hydrogen bond (p. 116), the effect of which on molar refraction is not known.²

Molar Refraction of Mixtures.—The experimental value of the molar refraction $[R]_{1,2}$ of a mixture of x_1 and x_2 mole fractions, respectively, of two substances of molecular weight M_1 and M_2 , is given by

$$[R]_{1,2} = \frac{n^2 - 1}{n^2 + 2} \cdot \frac{x_1 M_1 + x_2 M_2}{\rho}, \quad (7)$$

where n and ρ refer to the mixture. This quantity is found to be the sum of the contributions of the separate constituents, so that

$$[R]_{1,2} = x_1[R]_1 + x_2[R]_2. \quad (8)$$

where $[R]_1$ and $[R]_2$ are the individual molar refractions. Very few, if any, marked exceptions to this rule are known; it is particularly useful for the determination of the refractions of solid substances or of materials which are available only in small amounts. The molar refractivity of a solution of known composition is determined, by means of (7), from its measured refractive index and density. If the molar refraction of one of the constituents of the solution, e.g., the solvent,* is known, the value for the other can be readily calculated from (8). This procedure is frequently employed in connection with the determination of dipole moments (p. 549).

Measurement of Refractive Index.—The refractive index of a liquid is an important physical property, for in addition to its use as described above, it is frequently employed as a test of purity: it is determined by an instrument

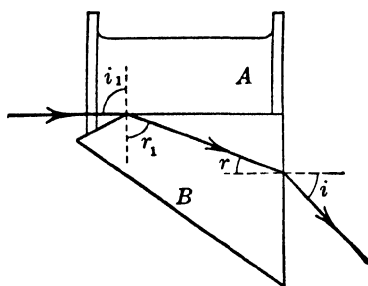


FIG. 114. Pulfrich refractometer

known as a **refractometer**. The Pulfrich type is convenient and capable of giving accurate results with little expenditure of time. The liquid to be examined is placed in a cylindrical glass cell *A* fixed to a right-angled prism *B* made of glass of high refractive index, generally about 1.6 (Fig. 114). Suppose a ray of monochromatic light, e.g., the sodium-*D* line, or one of the lines of the hydrogen spectrum, enters the prism parallel to its horizontal surface, i.e., at grazing incidence, and passes through as shown. By the laws of refraction $\sin i_1 / \sin r_1 = N/n$, where N is the refractive index of the glass prism and n that of the liquid; since i_1 is 90° , it follows that $\sin r_1 = n/N$. The direction at which the beam of light emerges into the air depends on the refractive index of the prism relative to that of air, which is unity; hence $\sin i / \sin r = N$, and so $\sin r = (\sin i)/N$. Since $r + r_1 = 90^\circ$, it follows that

$$\sin r = \cos r_1 = \sqrt{1 - \sin^2 r_1} \quad (9)$$

and substitution of $(\sin i)/N$ for $\sin r$ and n/N for $\sin r_1$, gives

$$n = \sqrt{N^2 - \sin^2 i}. \quad (10)$$

The Pulfrich refractometer permits of the measurement of the angle i , and if the index of refraction of the prism is known, that of the liquid can be calculated. Tables of $\sqrt{N^2 - \sin^2 i}$ for various values of i are supplied with the instrument, and so the refractive index n may be read off directly for the measured angle.

* The convention generally adopted in chemical literature is to use the suffix 1 for the solvent and 2 for the solute.

Refractive Dispersion.—The refractive index of a medium varies with the wave length of the light, and a relationship of the form

$$n = 1 + \sum \frac{a}{\nu_0^2 - \nu^2} \quad (11)$$

has been derived from both classical and quantum considerations. In (11), n is the refractive index measured with light of frequency ν ; a is a constant in each term of the series, being a measure of the strength of the oscillator of characteristic frequency ν_0 . The quantity ν_0 is the same as that in London's equation for intermolecular attraction (p. 299), and for simple substances it appears that $h\nu_0$ is almost identical with the ionization energy. For more complex molecules, such as liquid organic compounds, particularly if they contain a chromophoric group (p. 582), the quantity $h\nu_0$ probably represents the energy of an electronic transition, not necessarily ionization; the frequency ν_0 should thus correspond to that of a characteristic band in the spectrum of the molecule, and this is apparently true in a number of instances. If the frequencies in (11) are replaced by the reciprocals of the corresponding wave lengths, then it follows that

$$n = 1 + \sum \frac{a'\lambda^2}{\lambda^2 - \lambda_0^2}, \quad (12)$$

where a' is a constant which includes λ_0^2 . If at least one of the values of λ_0 is small, then that particular term becomes $a'\lambda^2/\lambda^2$, i.e., a' , and (12) reduces to

$$n = b + \sum \frac{a'\lambda^2}{\lambda^2 - \lambda_0^2}, \quad (13)$$

b being a constant, the summation now including only those terms for which λ_0 is not negligible. By squaring (13) and ignoring the terms involving λ^4 , it is seen that

$$n^2 = b^2 + \sum \frac{c\lambda^2}{\lambda^2 - \lambda_0^2}, \quad (14)$$

where c is another constant; this is identical in form with the refractive dispersion equation attributed to W. Sellmeier (1871), but which was apparently first derived by J. C. Maxwell (1869). By writing $\lambda^2/(\lambda^2 - \lambda_0^2)$ in the form $1/(1 - \lambda_0^2/\lambda^2)$, which is equal to $1 + \lambda_0^2/\lambda^2 + \lambda_0^4/\lambda^4 + \dots$, substitution in (14) gives

$$n^2 = b^2 + \sum c(1 + \lambda_0^2/\lambda^2 + \lambda_0^4/\lambda^4 + \dots); \quad (15)$$

$$\therefore n^2 = A + B/\lambda^2 + C/\lambda^4 + \dots, \quad (16)$$

which is similar to the formula deduced by A. L. Cauchy (1836).

Consider a simple form of (11), involving only one characteristic frequency; then

$$n = 1 + \frac{a}{\nu_0^2 - \nu^2}. \quad (17)$$

As long as the frequency ν of the light used for measuring the refractive index is considerably less than ν_0 , that is to say, at relatively long wave lengths, the refractive index changes relatively little with wave length; when ν approaches ν_0 , however, the refractive index begins to increase rapidly. This is the phenomenon of **anomalous refractive dispersion** which is observed in the vicinity of an absorption band of the molecule; it is responsible, to a large extent, for

much of the exceptional behavior, e.g., optical exaltation, observed in connection with molar refraction. Strictly speaking, the comparison of molar refractions should be made in a wave length region at a considerable distance from those in which anomalous dispersion becomes apparent (cf. p. 529). For saturated organic compounds the absorption bands controlling the dispersion are in the far ultraviolet portion of the spectrum, and hence measurements of refractivity made with the sodium-*D* line, or the α -, β - or γ -lines of hydrogen, are satisfactory, and the molar refraction is very closely an additive property. When the system contains a conjugated system of double bonds, however, the absorption band may appear in the visible region (see p. 582), and the measured refractive index may well be quite anomalous; the molar refraction is thus much higher than it should be normally, and the phenomenon of exaltation is apparent (cf. also p. 538).

The difference between the specific refractions for two wave lengths, e.g., the H_α and H_γ lines, is called the **specific dispersivity**; similarly, the **molar dispersivity** is defined as the difference between the molar refractions for two wave lengths, e.g., $[R]_\gamma - [R]_\alpha$. Attempts have been made to derive "dispersive constants" for atoms, but the results are unsatisfactory, for the dispersivity does not appear to be additive. When the observed molar dispersivities are compared with those calculated from the differences of the atomic equivalents for the corresponding wave lengths, as given in Table 82, anomalies, such as optical exaltation, are greatly magnified; this can be readily understood in the light of the discussion given above.

MOLAR POLARIZATION

Polarizability of Molecule.—The electrons and nuclei in any molecule are, to some extent, mobile and so when the molecule, whether it is polar or nonpolar, is placed in an electric field there will be a small displacement of the electrical centers with the result that a dipole, in addition to one which may already be present, will be induced in the molecule. If m is the electrical moment of the induced dipole produced by a field of intensity F acting on a single molecule, then

$$m = \alpha F. \quad (18)$$

The constant α is called the **polarizability** of the molecule; it is a measure of the ease with which the molecule can be polarized, that is the ease of displacement of positive and negative charges with respect to each other, in an electrical field. The polarizability, or deformability, as it is sometimes called, may vary in different directions if the molecule is not symmetrical; in that case α represents a mean value.

The Mosotti-Clausius Equation.—Suppose the strength of a uniform electric field produced by two charged plates is E_0 ; then in any nonpolar medium the field strength is reduced to E because the dipoles induced in the molecules act in opposition to the applied field. The ratio E_0/E is the dielectric constant D of the medium.* It can be shown by electro-

* The electrical capacity of the medium is altered in the inverse ratio, so that the dielectric constant is often defined as the ratio of the capacity of the condenser containing the given substance to that with a vacuum between the plates.

statics that

$$E_0 = E + 4\pi I \quad \text{or} \quad (D - 1)E = 4\pi I, \quad (19)$$

I being the induced electric moment per unit volume, i.e., $I = mn$, where n is the number of molecules per cc. The intensity of the electric field acting on each molecule, as required by (18), is F ; this is calculated by supposing a unit charge to be enclosed in a small spherical cavity which is relatively large as compared with a molecule, but small in comparison with the distance between the charged plates. The electrical intensity F is made up of several parts: there is first the field of the charges on the plates, i.e., E_0 ; then there is the force $-4\pi I$ due to the charges induced on the surfaces of the dielectric in contact with the plates, and $+\frac{4}{3}\pi I$ resulting from the charge induced on the surface of the spherical cavity; finally, the field caused by molecules within the cavity must be added. For a liquid or gas whose molecules are oriented at random in the absence of an external field the last contribution is zero, so that

$$F = E_0 + \frac{4}{3}\pi I - 4\pi I. \quad (20)$$

Introducing (19), it follows that

$$F = E + \frac{4}{3}\pi I = \left(\frac{D + 2}{3} \right) E. \quad (21)$$

By definition I is equal to mn , and with the aid of this relationship and (18) and (19), it can be readily seen that

$$\frac{D - 1}{D + 2} = \frac{4}{3}\pi n\alpha. \quad (22)$$

If ρ is the density of the medium between the charged plates, and M is its molecular weight, then the number of molecules n in unit volume is $N\rho/M$, where N is the Avogadro number. Substitution in (22) then results in

$$\frac{D - 1}{D + 2} \cdot \frac{M}{\rho} = \frac{4}{3}\pi N\alpha. \quad (23)$$

The left-hand side of this equation is given the symbol P and is called the **molar polarization** of the material; thus,

$$P = \frac{D - 1}{D + 2} \cdot \frac{M}{\rho}. \quad (24)$$

Since the applied field produces an induced charge in the molecule by the relative displacement, or distortion, of electrons and nuclei, P is often referred to as "induced" or "distortion" polarization.

The relationship between dielectric constant and polarizability, with M omitted, was first deduced by O. F. Mosotti (1850) and R. Clausius (1879), and the expression for the molar polarization, as written above, is generally known as the **Mosotti-Clausius equation**. It is now realized

that the assumption of a small spherical cavity containing a unit charge, on which the deduction is based, is only an approximation, and hence the equation cannot be regarded as exact; it is probably reasonably accurate, however, for a gas at low pressure. Other equations have been deduced which are more rigid, but they are too complicated to be considered here.

It will be seen from (24) that since D is a pure number, the molar polarization P is a volume; it is, in fact, a molar volume. Its value should be independent of temperature, and this has been found to be true in many cases, *provided the substance is nonpolar and the molecule has no permanent dipole moment*. It is for such molecules that the relationship derived by J. C. Maxwell (1881) connecting the dielectric constant of a substance and its refractive index for light of long wave length, i.e.,

$$D = n_{\infty}^2, \quad (25)$$

is applicable, so that substitution in (24) gives

$$P = \frac{n_{\infty}^2 - 1}{n_{\infty}^2 + 2} \cdot \frac{M}{\rho} = \frac{4}{3} \pi N \alpha. \quad (26)$$

The middle term in this equation is the molar refraction for radiations of long wave length [cf. equation (6)], and this quantity is evidently a measure of the polarizability of the nonpolar molecule in an electric field. As indicated on p. 529, the value of n_{∞} cannot be obtained by extrapolation to infinite wave length of refractive indices measured with visible light; visible light can only displace electrons while the positively charged nuclei remain unaffected. That is to say, the polarizability or deformability α may be divided into two parts, viz., α_E and α_A , called the electron and atom (nuclear) polarizabilities, respectively; the former *only* is measured by visible light and extrapolation of α_E values can never give α , for this includes α_A . If n is the refractive index of visible light,* then

$$\frac{n^2 - 1}{n^2 + 2} \cdot \frac{M}{\rho} = \frac{4}{3} \pi N \alpha_E = P_E, \quad (27)$$

the quantity P_E obtained in this way being called the **electron polarization**. It is equal to the molar refraction for visible light, and represents that part of the total induced polarization P which is due to the distortion or deformation of the electronic shells in the molecule. To obtain the *whole* of P it is, therefore, necessary actually to measure the refractive index with radiations of long wave length, e.g., infrared rays, for vibrations of low frequency are able to displace the relatively heavy atomic nuclei as well as the electrons; in other words, both α_E and α_A are included. The difference between the total induced or distortion polarization P , as given by (26), and the electron polarization P_E is called the **atom**

* The measurement must, of course, not be made in or near an absorption band where the behavior is anomalous.

polarization P_A ; that is,

$$P = \frac{n_{\infty}^2 - 1}{n_{\infty}^2 + 2} \cdot \frac{M}{\rho} = P_E + P_A. \quad (28)$$

Measurements of refractive index with infrared radiations are difficult, and values of P_A are generally determined indirectly (p. 548); for most substances they are small, not more than 2 or 3 cc., and as a rough approximation they may be taken as 5 to 10 per cent of P_E for the D -line of sodium. In a few exceptional cases, e.g., *p*-benzoquinone, *p*-dinitrobenzene and metal acetylacetonates, the atom polarization is exceptionally large, for reasons which are not yet clear.

Refraction of Bonds and Electron Groups.—The molar refraction for visible light varies with the wave length, and so the electron polarization is somewhat uncertain; for most purposes, however, the D -line of sodium may be used, so that (27) becomes

$$\frac{n_D^2 - 1}{n_D^2 + 2} \cdot \frac{M}{\rho} = [R]_D = \frac{4}{3}\pi N\alpha_E. \quad (29)$$

The molar refraction is thus a measure of the deformability of the electronic sheaths, and hence it would appear that instead of dissecting molar refractions into the atomic equivalents (p. 529), a more rational division could be made from the standpoint of the electron groups concerned. Such a treatment of refraction data was first proposed by A. L. von Steiger (1921), and was developed independently by K. Fajans (1924) and C. P. Smyth (1925). Since the four valences in methane can be regarded as equivalent, the eight valence electrons may be assumed to be uniformly distributed between the four C—H bonds; the refraction of the pair of electrons shared between carbon and hydrogen, or, in other words, the refraction equivalent of the C—H bond, is thus one-fourth of the molar refraction of methane; that is,

$$R_{C-H} = \frac{1}{4}R_{CH_4} = \frac{1}{4}R_C + R_H, \quad (30)$$

where R_C and R_H are the ordinary refraction equivalents of carbon and hydrogen as given in Table 82 (p. 530). Similarly the refraction of the two electrons joining two singly-linked carbon atoms, that is of the C—C bond, is obtained by subtracting the value for six C—H linkages, as obtained above, from the molar refraction of ethane; thus,

$$\begin{aligned} R_{C-C} &= R_{C_2H_6} - 6R_{C-H} \\ &= (2R_C + 6R_H) - 6(\frac{1}{4}R_C + R_H) = \frac{1}{2}R_C, \end{aligned} \quad (31)$$

so that the refraction of the electron pair of the C—C bond is half the refraction equivalent generally ascribed to a carbon atom. The refraction of the four electrons constituting the double bond in the C=C grouping is given by

$$R_{C=C} = R_{C_2H_4} - 4R_{C-H} \quad (32)$$

and this is readily shown to be equal to the sum of the equivalents of a carbon atom and a double bond. In an analogous manner $R_{C\equiv C}$ can be derived by adding the usual triple-bond equivalent to $\frac{3}{2}R_C$.

If the refraction of three C—H bonds is subtracted from that of methyl chloride, the remainder is the refraction of the eight electrons of the chlorine

atom, two of which are shared with a carbon atom; the same result is obtained by taking one-fourth of the molar refraction of carbon tetrachloride. The refractions of the other halogens, when attached to carbon, may be determined in an analogous manner. Values for the electron octets of halogens when one pair is shared with boron, silicon or sulfur may be calculated from the molar refractions of the corresponding halides, e.g., BCl_3 , SiCl_4 and SF_6 . Some of the results obtained for the refractions of various electron groups, for the sodium *D*-line, are given in Table 83; the dots in the formulae represent electrons (cf.

TABLE 83. REFRACTION OF ELECTRON GROUPS (*D*-LINE)

Electron Group	Refrac- tion	Electron Group	Refrac- tion	Electron Group	Refrac- tion
C:H	1.705	C: $\ddot{\text{F}}$:	1.60	Si: $\ddot{\text{Cl}}$:	7.04
C:C	1.209	C: $\ddot{\text{Cl}}$:	6.57	S: $\ddot{\text{F}}$:	1.95
C::C	4.15	C: $\ddot{\text{Br}}$:	9.47	Se: $\ddot{\text{F}}$:	2.23
C:::C	6.025	C: $\ddot{\text{I}}$:	14.51	Te: $\ddot{\text{F}}$:	2.47

p. 98). Since the refraction may be regarded as a measure of the electron polarizability, it follows that the figures in the table give a quantitative indication of the polarizability or deformability of the electron groups in an electric field. The values hold not only for external fields but are also applicable when the fields result from the presence of strongly polar groups within the molecule; this is of importance in connection with modern theories of reactivity in organic compounds. It is seen that the polarizabilities of the electron octets in the halogen series increase markedly with increasing atomic weight, and that for any given halogen the ease of deformation or distortion depends on the other atom with which it shares a pair of electrons. The polarizability of a double bond, even for each electron, is evidently much greater than of a single bond; it is for this reason that electrical effects are more readily transmitted along an unsaturated than through a saturated carbon chain. The optical exaltation effect (p. 530) accompanying conjugated double bonds may be, to some extent, an indication of increased polarizability of the electron system.

The study of the refraction or polarizability of electron groups provides a physical basis for the necessity of attributing refraction equivalents to double and triple bonds; it also offers an interpretation of the fact that the refraction equivalent of carbon is constant whereas with oxygen and nitrogen it varies with the type of compound. The four electron pairs associated with quadri-valent carbon are probably almost identical, and substitution of one atom for another attached to one of these electron pairs does not affect appreciably the polarizability of the remainder. In elements like oxygen, nitrogen, sulfur, phosphorus, etc., possessing lone pairs of electrons in the combined state, it is not surprising that the polarizability of the octet varies according to the number and nature of the attached atoms.³

Refraction and Deformation of Ions.—The net positive charge of a cation should have the effect of binding the electrons tightly to the

nucleus, whereas a loosening tendency should exist in anions (cf. p. 384); it follows, therefore, that of two ions having the same inert-gas structure the cation should be less polarizable than the anion. Further, the polarizability of any ion of a given series, e.g., alkali metals or halogens, would be expected to increase with increasing size. These results may be expressed quantitatively in the form of the relationship (M. Born and W. Heisenberg, 1924)

$$\alpha = c/(Z - y)^3, \quad (33)$$

where Z is the atomic number of the element, c is a constant for a series of ions with the same inert-gas structure, and y has the values 0, 6, 13, 28 or 46, according as the structure is that of helium, neon, argon, krypton or xenon, respectively. From these considerations the following general conclusions may be reached concerning the relative refractions of various ions and inert gases:

$$R_{F^-} > R_{Ne} > R_{Na^+} \quad \text{and} \quad R_{Cl^-} > R_A > R_{K^+},$$

and so on for the other ions having the electronic structures of krypton and xenon, and

$$R_{He}/R_{Li^+} > R_{Ne}/R_{Na^+} > R_A/R_{K^+} > R_{Kr}/R_{Rb^+} > R_{Xe}/R_{Cs^+}.$$

The molar refractions of all the alkali halides have been measured in the solid state and in dilute aqueous solution. The results do not differ greatly, but such discrepancies as do arise can be accounted for by the closer proximity of the ions in the crystal, and by the mutual influence of ions and water molecules in solution. A positive ion will tend to attract the electrons of a solvent molecule, and by causing a general tightening will decrease its polarizability; a negative ion, on the other hand, will tend to have the opposite effect. The influence of an ion on the refraction of water will thus depend on the sign of the former, and also on its size because this will determine the strength of its electrical field. The smallest ions, lithium and fluorine, will consequently have an appreciable effect, whereas that for larger ions is probably very small. Bearing all these points in mind, including the relative magnitudes of the ionic refractions noted above, and assuming the molar refraction of a salt is approximately the sum of the values for its constituent ions, K. Fajans and G. Joos (1924), following J. A. Wasastjerna (1923), drew up a table of approximate refractions of "gaseous ions," that is of ions assumed to be so far removed from neighboring ions and from solvent molecules as to have their polarizabilities unaffected. The values for the alkali metals and halogen ions, with some slight modification, are given in Table 84. The refraction of the corresponding inert gas is recorded for comparison. It is of interest to compare the refractions of the halogen ions with those of the corresponding atoms attached to carbon (Table 83); the loosening of the electronic shell accounts for the larger deformability of the former.

TABLE 84. REFRACTIONS OF "GASEOUS IONS"

Cations		Inert Gas		Anions	
Li ⁺	0.2	He	0.5	—	—
Na ⁺	0.5	Ne	1.00	F ⁻	2.5
K ⁺	2.1	A	4.20	Cl ⁻	9.0
Rb ⁺	3.5	Kr	6.37	Br ⁻	12.5
Cs ⁺	6.1	Xe	10.42	I ⁻	19.0

A halogen hydride has the same number of electrons as the halide ion itself; consequently, apart from any effects resulting from the neutralization of the charge, the refraction should remain unchanged. In actual fact, as might be expected, the hydride has the smaller value and the discrepancy increases with increasing polarizability of the anion, as the results in Table 85 show. The addition of a proton to a halide ion thus

TABLE 85. COMPARISON OF REFRACTIONS OF HALOGEN HYDRIDES AND HALIDE IONS

	F	Cl	Br	I
R_{X^-}	2.5	9.0	12.5	19.0
R_{HX}	1.9?	6.67	9.14	13.74
Difference	0.6?	2.3	3.4	5.3

results in a decrease of polarizability, which corresponds to a general tightening up of the electronic shells. This type of deformation is evidently equivalent to the conversion of an electrovalent into a covalent linkage, for the bond between hydrogen and halogen in the gaseous hydrogen halides is mainly covalent. It will be seen, too, that the values of R_{HX} in Table 85 are very similar to those in Table 83 for halogens bound by a covalence to a carbon atom.

A comparison of the molar refractions of the solid alkali halides with the sums of the refractions of the appropriate gaseous ions leads to interesting conclusions; the differences are recorded graphically in Fig. 115.* The results can be explained in the following manner: the positive electric field of a cation causes a tightening of the electronic sheaths of the anion, with a consequent decrease of polarizability and refraction, whereas the negative field of the anion tends to loosen the electrons of the cation so that the refraction increases. As indicated above, anions are more polarizable than cations, while the latter being smaller, for the same structure, exert the more intense field, that is, they have the greater polarizing power; hence the deformation leading to a decrease of polarization is likely to be more marked than the reverse effect. That this is the case is apparent from Fig. 115; the deformation effect, which is equivalent to the change from an electrovalence to a covalence or, in other words, to an increased contribution of nonpolar character (p. 109), is more marked the larger, that is, the more polarizable, the anion and the smaller the cation, and hence the greater its polarizing effect. For such substances,

* The data for the cesium halides are not included as the crystal structure of the chloride, bromide and iodide differs from that of the other alkali halides (p. 386).

e.g., lithium iodide, the polarization of the cation by the anion, leading to an increase of refraction, is quite negligible. Not only is the cation small and almost nonpolarizable, but the polarizing effect, or field strength, of the large ion is very small. The only possibility for the detection of this positive influence is when a large, appreciably polarizable, cation is associated with a small anion, of small polarizability and relatively high external field, as in rubidium and potassium fluorides.

The views developed here have been found to be applicable in many other instances; anions are always more deformable than the corresponding cations, and the polarizing or deforming power of the latter increases

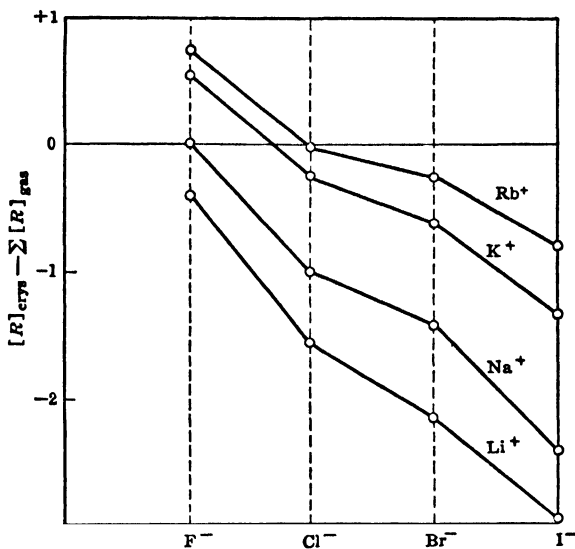


FIG. 115. Refractions of crystalline and gaseous ions

with increasing charge and decreasing size. The cations of heavy metals, e.g., Ag^+ , Cd^{++} and Hg^{++} , which do not have the inert-gas structure, have been found to exert a stronger polarizing effect than ions of the inert gas type of similar size and charge. The unusual properties of the halides, and particularly the iodides, of such metals, e.g., low conductance of the mercuric salts in solution, formation of complex ions, the color of the iodides, etc., are explicable by the deformation of the anions which results in a tendency for the valence bond to become more covalent in character.⁴

Polarizability and Molecular Radius.—It has been deduced by electrostatics that the moment m induced in a perfectly conducting sphere of radius r when placed in an electric field of strength F is given by

$$m = r^3 F, \quad (34)$$

and comparison with (18) shows that for a spherical molecule, assumed to

be a perfect conductor, the polarizability is equal to the cube of the radius; thus,

$$\alpha = r^3. \quad (35)$$

It follows, therefore, from (26) that

$$\frac{n_{\infty}^2 - 1}{n_{\infty}^2 + 2} \cdot \frac{M}{\rho} = \frac{4}{3}\pi N\alpha = \frac{4}{3}\pi N r^3, \quad (36)$$

so that the molar refraction is equal to the actual volume of the molecules in 1 mole. Strictly the refractive index for long wave lengths n_{∞} is involved, but for approximate purposes data for visible light are satisfactory. With the aid of (36) an estimate of molecular radius may be made from the refraction; the results are only reliable for simple molecules which may be regarded as approximately spherical. In such cases, e.g., hydrogen, nitrogen, chlorine, argon, methane, water and methyl alcohol, the molecular radii calculated in this manner are very similar to those obtained from the van der Waals constant (p. 294) or from viscosity measurements (p. 282). When the substance is normally a liquid, as in the last two instances mentioned, the measurements of refractive index and density are more conveniently made in that form; as already seen, the molar refraction is almost the same for gaseous and liquid states. The refractive indices of gases do not differ greatly from unity and considerable care has to be taken to obtain accurate results; the calculation of the molar refraction may, however, be simplified by utilizing the fact that $n^2 + 2 \approx 3$, so that for a gas

$$\alpha = r^3 = \frac{n^2 - 1}{4\pi N} \cdot \frac{M}{\rho}. \quad (37)$$

Some of the molecular radii obtained from refraction data are quoted in Table 86: the last column gives the radii (r_b) evaluated from the van der

TABLE 86. MOLECULAR RADII FROM MOLECULAR REFRACTIONS

	$(n^2 - 1)$	α	r	r_b
Hydrogen	0.272×10^{-3}	0.80×10^{-24}	0.93 Å.	1.29 Å.
Argon	0.560	1.64	1.18	1.46
Chlorine	1.436	4.25	1.62	1.60
Nitrogen	0.581	1.73	1.20	1.48
Water	0.51	1.48	1.14	1.08
Methyl alcohol	1.20	3.37	1.50	1.45

Waals constant. It is of special interest that the first reliable values for the radii of ions in crystals were obtained by J. A. Wasastjerna (1923) from ionic refractions, after making corrections for the approximations involved.

For reasons which will be apparent shortly, (37) is applicable to polar as well as to nonpolar substances, as is evident from the inclusion of water and methyl alcohol, both of which are polar, in Table 86. The

corresponding equation (23), involving the dielectric constant instead of the refractive index, viz.,

$$\frac{D-1}{D+2} \cdot \frac{M}{\rho} = \frac{4}{3} \pi N \alpha = \frac{4}{3} \pi N r^3, \quad (38)$$

can be used to determine radii of nonpolar molecules only; for a substance possessing a permanent electric moment the Maxwell relationship, $D = n_{\infty}^2$, no longer holds.

DIPOLE MOMENTS

Polar Molecules.—For polar compounds there arises, in addition to the discrepancy noted above, the difficulty that the molar polarization P , as given by the Mosotti-Clausius equation

$$P = \frac{D-1}{D+2} \cdot \frac{M}{\rho}, \quad (24a)$$

is not independent of temperature, especially for vapors and dilute solutions; the explanation* of this phenomenon, and the important concept of **dipole moments** arising from it, is due to P. Debye (1912). In the derivation of Mosotti and Clausius it is assumed that before the application of an electric field there is a uniform distribution of the charges in the molecule, so that it has no permanent moment; it is possible, however, that the charges are so distributed as to give rise to an electric doublet or dipole, even in the absence of an applied field, and the molecule then has a permanent dipole moment (p. 108). In the measurement of dielectric constant an oscillating electric field of frequency about 10^6 to 10^7 cycles per sec. is usually employed; the period of alternation is then greater than the time of relaxation of the polar molecule, that is, the time taken for the permanent dipole to orient itself in the electric field. With visible or infrared radiation, however, the oscillations are so rapid that there is only sufficient time for the electrons and nuclei to undergo deformation; the presence of the permanent moment has, therefore, no influence on the refractive index. It follows then that even for polar molecules the molar refraction should be a definite constant, as found in practice, even though the molar polarization derived from the dielectric constant varies with temperature.

The potential energy U of a molecule, possessing a permanent moment μ , when placed in an electric field of intensity F (cf. p. 534), so that the axis of the dipole makes an angle θ with the direction of the field (Fig. 116), is given by

$$U = -\mu F \cos \theta \quad (39)$$

and the permanent moment in the direction of the field is $\mu \cos \theta$. For the present, the effect of the additional dipole produced by induction is neglected.

* A virtually correct interpretation of the behavior of molecules possessing a permanent electric moment was given by J. A. Fleming (1900).

According to the Maxwell-Boltzmann distribution law (cf. p. 270) the number of molecules having their dipole axes within a solid angle $d\Omega$ at any instant is $Ae^{-U/kT}d\Omega$, where A is determined by the total number of molecules present, and k is the Boltzmann constant (p. 264). The total number of molecules can be obtained by integrating $Ae^{-U/kT}d\Omega$ over all directions in space, i.e., for solid angles between 0 and 2π ; thus,

$$\text{Total number of molecules} = \int_0^{2\pi} Ae^{-U/kT}d\Omega = \int_0^{2\pi} Ae^{\mu F \cos \theta / kT} d\Omega. \quad (40)$$

Since each molecule has a moment $\mu \cos \theta$ in the direction of the field, the total moment within the solid angle $d\Omega$ is equal to $Ae^{-U/kT}\mu \cos \theta d\Omega$, and the resultant value for all the molecules can again be obtained by integration between 0 and 2π . The *mean* moment \bar{m} of one molecule in the direction of the field,

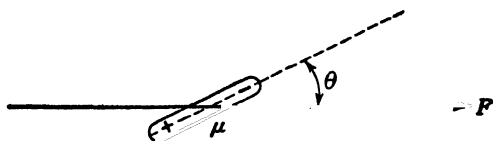


FIG. 116. Dipole in electric field

can then be derived by dividing the total moment by the number of molecules; that is,

$$\bar{m} = \frac{\int_0^{2\pi} Ae^{\mu F \cos \theta / kT} \mu \cos \theta d\Omega}{\int_0^{2\pi} Ae^{\mu F \cos \theta / kT} d\Omega}. \quad (41)$$

By writing x for $\mu F/kT$, it can be shown that

$$\frac{\bar{m}}{\mu} = \frac{e^x + e^{-x}}{e^x - e^{-x}} - \frac{1}{x} = \coth x - \frac{1}{x} = L(x), \quad (42)$$

where $L(x)$ represents the function derived by P. Langevin (1905) in connection with his work on the paramagnetism of gaseous molecules (p. 615). If x is small, particularly for small field strengths, when saturation effects do not arise, it is found that

$$L(x) \approx \frac{1}{3}x \quad (43)$$

and (42) then gives

$$\bar{m} = \frac{1}{3}\mu x = \frac{\mu^2}{3kT} F \quad (44)$$

It is now necessary to add to this the moment αF produced by induction (p. 534) in order to obtain the total moment per molecule in the direction of the electric field; that is,

$$m = \alpha F + \frac{\mu^2}{3kT} F. \quad (45)$$

It follows from (18), (23) and (24) that $P = \frac{1}{3}\pi N m/F$, and this same relationship can be derived for a polar molecule by the method given on p. 535; hence

from (45)

$$P = \frac{D-1}{D+2} \cdot \frac{M}{\rho} = \frac{4}{3}\pi N\alpha + \frac{4}{3}\pi N \left(\frac{\mu^2}{3kT} \right), \quad (46)$$

which is known as the **Debye equation**. The above derivation is not entirely free from objection, but the same result has been obtained by Debye and others using more rigid methods based on quantum mechanics, and so it may be accepted as correct.⁵

The total polarization (P) of a molecule possessing a permanent dipole moment is seen by (46) to be the sum of two terms: the first gives the polarization resulting from induction or deformation (p. 536), that is, the distortion polarization (P_D), whereas the second, due to the permanent moment possessed by the molecule, is called the **orientation polarization** (P_O). The distortion polarization is the sum of the electron (P_E) and atom (P_A) polarizations (see p. 537), so that

$$\frac{D-1}{D+2} \cdot \frac{M}{\rho} = P = P_D + P_O; \quad (47)$$

$$\therefore P_D = P_E + P_A = \frac{4}{3}\pi N\alpha \quad \text{and} \quad P_O = \frac{4}{3}\pi N \left(\frac{\mu^2}{3kT} \right). \quad (48)$$

The term orientation polarization arises from the fact that the electric doublets are oriented by the applied field so that the positive ends of the permanent dipoles point towards the negatively charged plate; the intensity of the field within the medium is thus diminished as compared with that in a vacuum, and hence the dielectric constant exceeds unity by an appreciable amount. Substances with permanent moments thus generally have high dielectric constants. If the molecules containing the electric doublets were fixed and unable to be oriented in the field the orientation polarization would be zero; this is true for substances in the solid state provided free rotation is not possible (p. 422). It is important to realize, therefore, that (46) can hold only if there is no restriction whatever to the movement of the dipoles in an electric field; even in the liquid state molecules are not quite free from mutual interaction and so the Debye equation should be applied only to gases and vapors.

Measurement of Dielectric Constant.—The determination of dielectric constant involves the measurement of the capacity of a condenser filled with the substance under investigation, which is compared with that for a vacuum between the plates; the method most commonly employed, on account of simplicity, is the **heterodyne beat method**. Two similar electron-tube oscillators, containing suitable capacities and inductances to give a frequency of about 10^6 cycles per sec., are coupled to an amplifier. One of the oscillators contains fixed condensers and hence has a fixed frequency, whereas the other has an accurately calibrated standard variable condenser in parallel with the experimental cell; the latter may consist of a number of parallel flat metal plates or of two, or more, concentric metal, or metal-coated, cylinders capable of acting as a condenser. By adjustment of the variable condenser the frequency

of the oscillator can be altered so as to coincide with that in the fixed oscillator: no sound can then be detected in the amplifier. More accurate results can be obtained by listening for a beat note of definite frequency, which is equal to the frequency difference between the two oscillators. If the readings of the standard condenser are taken, with the experimental cell in and out of the circuit, the capacity can be determined, both empty and full of the substance to be examined; the dielectric constant can then be evaluated. In the **resonance method**, which is also often employed, a fixed frequency oscillator, generally controlled by a quartz crystal, is coupled to a "resonance" circuit containing inductance and capacity, the latter in the form of a standard variable condenser and the experimental cell. At a given frequency, determined by the inductance and capacity, resonance occurs between the two systems, and this is indicated by the maximum deflection of a galvanometer associated with the resonance circuit. The readings of the standard condenser are then taken, with and without the experimental cell, and the dielectric constant calculated as indicated above. In practice various corrections must be applied with both methods, e.g., for the capacity of leads; this is generally done by making measurements on a standard substance of known dielectric constant, e.g., air or carbon dioxide for gases, or benzene for liquids.⁶

Polarization and Temperature.—As the temperature is raised the thermal movement of the molecules increases and this tends to oppose the orienting effect of the applied electric field on the permanent molecular dipoles; the orientation polarization should thus decrease with increasing temperature, in agreement with (46). The dependence of the total polarization on temperature may be seen by writing the Debye equation (46) in the form

$$P = a + \frac{b}{T}, \quad (49)$$

where a and b are constants for any given substance; thus,

$$a = \frac{4}{3}\pi N\alpha = P_A + P_E \quad (50)$$

and

$$b = \frac{4}{3}\pi N \left(\frac{\mu^2}{3k} \right) = P_O T. \quad (51)$$

According to (49), therefore, the plot of the total molar polarization of a gas or vapor, calculated from the dielectric constant, against the reciprocal of the absolute temperature should be a straight line. If the substance is nonpolar then μ , and hence b , is zero and the line will be parallel to the $1/T$ axis, but if it is polar the slope, equal to b , is determined by μ , the permanent dipole moment of the molecule. The experimental data for carbon tetrachloride, hydrogen chloride and methyl chloride are recorded in Fig. 117; the results are in agreement with expectation, for the first of these is symmetrical and nonpolar, whereas the others are polar molecules.

Determination of Dipole Moments: (i) **Vapor-Temperature Method.**—The most satisfactory method, although involving experimental diffi-

culties, for the determination of dipole moments is to measure the dielectric constant and density of the vapor at a series of temperatures. If the substance to be studied decomposes on heating, then observations can be made under reduced pressures. The total polarization is then calculated by (47) and plotted against $1/T$, as in Fig. 117; the value of b , the

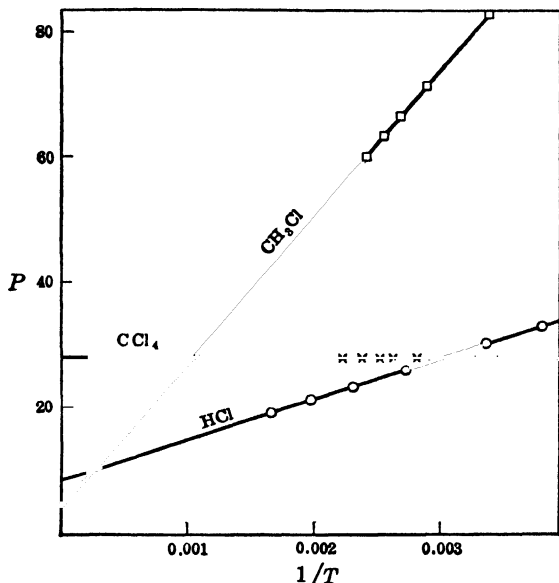


FIG. 117. Variation of total polarization with temperature

slope of the straight line, is then measured and from (51), it is seen that

$$\mu = \sqrt{9bk/4\pi N}. \quad (52)$$

Since N is 6.02×10^{23} and k is 1.38×10^{-16} erg per degree, it can be readily shown that

$$\mu = 0.0128 \sqrt{b} \times 10^{-18}, \quad (53)$$

so that the dipole moment can be calculated. If the density is expressed in g. per cc., then since the dielectric constant is a pure number, the polarization, like the molar refraction (p. 529), is given in cc. by (47). The dipole moment is the product of a charge and a distance, and it can be shown that if P is in cc. the value of μ as determined by (53) is in e.s. unit \times cm. It is found in practice that b is generally about 10^2 to 10^4 , so that dipole moments are of the order of 10^{-18} e.s. unit \times cm.; this unit is called a **Debye**, and given the symbol D , so that (53) may be written

$$\mu = 0.0128 \sqrt{b} D. \quad (54)$$

It is to be expected that all moments should be of the order of 10^{-18} e.s.-cm. unit; the unit charge of electricity is 4.8×10^{-10} e.s. unit and molecular dimensions are about 10^{-8} cm., and the product, giving the approximate magnitude of molecular dipole moments, must be about 10^{-18} .

It may be noted, incidentally, that the plot of the total polarization against $1/T$ provides a method for the evaluation of the atom polarization. According to (50) the intercept on the P axis, that is α , is equal to $P_A + P_E$, and since P_E is almost identical with the molar refraction for visible light, it is possible to evaluate P_A . It is from measurements of this kind that the atom polarization is known to be small except for a few instances (see p. 537).

(ii) **Refraction Method.**—By writing (47) and (48) in the form

$$P = P_D + \frac{4}{3}\pi N \left(\frac{\mu^2}{3kT} \right), \quad (55)$$

it follows from (51) and (53) that

$$\mu = 0.0128 \sqrt{(P - P_D)T}, \quad (56)$$

where P_D is equal to $P_A + P_E$. If P is determined by measuring the dielectric constant of the vapor at any one temperature T , and P_D is obtained from the refractive index with infrared radiation, μ can be evaluated. This method is simpler than the previous one, since it involves measurement at only one temperature. The value of P_D is, however, by no means easy to obtain, and so the molar refraction for the D -line is taken as P_E and to it is added 5 per cent as an approximate allowance for P_A , and the total assumed to be P_D . The distortion polarization can sometimes be obtained from the dielectric constant of the solid; as seen above, provided there is no rotation, the orientation polarization is zero or very small, and the total polarization is about the same as P_D . The method of evaluating dipole moments by means of (56) is satisfactory provided $P - P_D$ is relatively large; if it is small, when μ is about 0.5 d. or less, the error in the determination of P_D leads to seriously incorrect values of the dipole moment.

(iii) **Dilute Solution Method.**—When a polar substance is dissolved in a nonpolar solvent the total polarization of the solution, if dilute, is the sum of the polarizations of the constituents. The total polarization $P_{1,2}$ of the mixture is given by

$$\frac{D - 1}{D + 2} \cdot \frac{x_1 M_1 + x_2 M_2}{\rho} = P_{1,2}, \quad (57)$$

where D is the measured dielectric constant and ρ the density of the solution; x_1 and x_2 are of the mole fractions of the nonpolar solvent and polar solute, respectively, whose molecular weights are M_1 and M_2 . Since $P_{1,2}$ is the sum of the polarization contributions of the two components, it follows that

$$P_{1,2} = x_1 P_1 + x_2 P_2, \quad (58)$$

where P_1 and P_2 are the molar polarizations of the nonpolar and polar substances, respectively. Since P_1 and $P_{1,2}$ can be evaluated from measurements of dielectric constant and density of solvent and dilute solution, P_2 can be calculated from the known composition of the solution, expressed in mole fractions. An indication of the type of results obtained in this manner is given by the data in Table 87, for dilute solutions of

TABLE 87. MOLAR POLARIZATION OF ETHYL ETHER FROM MEASUREMENTS IN SOLUTION

x_1	x_2	D	ρ	P_2
1.00000	0.00000	2.033	0.7784	($P_1 = 27.65$)
0.95280	0.04720	2.109	0.7751	58.11
0.91146	0.08854	2.178	0.7720	58.17
0.87675	0.12325	2.246	0.7691	57.53
0.82690	0.17310	2.317	0.7664	57.17

diethyl ether in cyclohexane at 20° c. The total polarization P_2 of the solute varies to some extent with the concentration of the solution, for reasons to be considered shortly, and for the calculation of the dipole moment the results are extrapolated either graphically, or by means of suitable equations, to infinite dilution. Since the total polarization of the dissolved substance is known, the dipole moment may be obtained by means of (56), for the dilute solution is assumed to be equivalent to a gas as far as freedom of orientation of the dipole molecules is concerned. The distortion polarization is assumed, as before, to be equal to the molar refraction for the D -line, together with an allowance of 5 per cent for the atom polarization. The value of $[R]_D$ can be calculated (p. 532) from measurements of the refractive index of the solutions.

Approximate extrapolation of the results in Table 87 gives a value of about 58.5 for P_2 for diethyl ether at infinite dilution at 20°; the molar refraction is 22.48, and hence P_D is approximately 23.5. It follows, therefore, from (56) that the dipole moment of the ether is 1.29 D.

Although the method just described has been very largely used for the determination of dipole moments, there is some doubt whether (57) is applicable to a solution of a polar substance in a nonpolar solvent (W. H. Rodebush, 1940). However, by the use of a more accurate relationship between the dielectric constant and the polarization, it is possible to use measurements on solutions to calculate the dipole moments of various solutes. Most of the data to be found in the literature, however, are based on the procedure explained above.⁷

Influence of Solvent.—It was originally considered that dipole moments derived from measurements on solutions, extrapolated to infinite dilution, were independent of the nature of the solvent and equal to the value for the vapor, but it is now realized that this view is incorrect. The measured polarization in solution is governed by the dielectric constant of the medium, and for the majority of substances the polarization, and hence the dipole moment, decreases with increasing dielectric constant. The solvent effect is generally attributed to the polarization induced in the molecules of solvent by the electric fields of the molecular dipoles of the solute. The extent of this induced polari-

zation will depend on the polarizability of the solvent and this is determined by the dielectric constant [cf. equation (22)]; hence the solvent effect is a function of the latter, and the relationship

$$\frac{\mu_{\text{sol}}}{\mu_{\text{vap.}}} = 1 + A \frac{D - 1}{D + 2} \quad (59)$$

has a theoretical basis. In this equation $\mu_{\text{sol.}}$ is the dipole moment in a solution of dielectric constant D , and $\mu_{\text{vap.}}$ is that of the vapor, which is the correct value measured in a medium of dielectric constant of almost unity. The sign and magnitude of the constant A depend on the shape of the molecule and the position of the dipole in it (K. Higasi, 1936). As a general rule, if the dipole lies parallel to the length of the molecule, A is positive, and the dipole moment decreases with increasing dielectric constant of the solvent; this is true for most substances, which thus exhibit a "negative" solvent effect. When the direction of the dipole is perpendicular to the longer axis of the molecule, e.g., chloroform, the solvent effect is "positive," that is, polarization and dipole moment increase with the dielectric constant. It should be noted that according to the theory of the Kerr effect (p. 557), molecules of the second type should have negative Kerr constants, and this appears to be generally true.

The value of A in (59), derived from molecular dimensions, can be determined only in the very simplest cases, and so it is rarely possible to calculate with any certainty the true dipole moment $\mu_{\text{vap.}}$ from measurements in solution. For substances with moments less than 2 D, the difference between $\mu_{\text{vap.}}$ and $\mu_{\text{sol.}}$ in a solvent of low dielectric constant, e.g., hexane ($D = 1.91$) or cyclohexane ($D = 2.02$), is relatively small and may be neglected. When the dipole moment is large, (59) shows that the discrepancy may be appreciable; for example, the dipole moment of nitrobenzene, obtained from measurements in benzene solution, was for many years believed to be 3.9 D, but the correct value is now taken as 4.23 D.

The realization that the measured polarization of a solute depends on the dielectric constant of the solution has helped to clear up a number of difficulties. For instance, it accounts for the decrease of P_2 frequently observed with increasing concentration, as in Table 87. The view was held at one time that "dipole association" was responsible, but it now appears that the results can be explained adequately without this assumption. An outstanding example of the variation of polarization of the solute was that of nitrobenzene in benzene solution; P_2 was found to decrease from 334 cc. at infinite dilution to 94 cc. in pure nitrobenzene. It was suggested that there was considerable dipole associa-

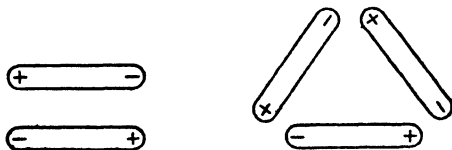


FIG. 118. Dipole association

tion, which would be accompanied by a decrease of polarization because of the mutual compensation of the dipoles (Fig. 118), the extent increasing with increasing concentration. It must be remembered, however, that the dielectric constant of benzene is about 2.2 as compared with 36 for nitrobenzene, and this

change will inevitably be accompanied by a marked decrease of polarization. There is a possibility that dipole association does occur in pure liquids and in concentrated solutions, but the amount is probably small; * it could only be estimated after making allowance for the expected variation of polarization with the dielectric constant of the medium. It is important to emphasize that when the solute is a substance containing an $-OH$ group, and so is definitely associated (p. 115), the extent of association will vary with the concentration, and will consequently affect the measured polarization; this type of association must be clearly distinguished from dipole association, which involve forces of a weaker character.

It would be imagined, at first sight, that the accurate method for determining dipole moments of vapors, by plotting the polarization against the reciprocal of the absolute temperature, could be applied to dilute solutions; it has been found by experience, however, that the results are incorrect. The reason is that the dielectric constant of the medium decreases with increasing temperature; the solvent effect thus varies with temperature and this is superimposed on the normal variation of the total polarization. According to (51) the product P_0T should be independent of temperature; this is true for gases and vapors, but not for solutions because of the varying solvent influence.^a

Results.—The dipole moments of many inorganic and organic compounds have been measured; a few of the results are given in Table 88,

TABLE 88. DIPOLE MOMENTS OF MOLECULES

Substance	μ	Substance	μ	Substance	μ
INORGANIC COMPOUNDS.					
Hydrogen chloride	1.03 D.	Ammonia	1.46 D.	Hydrogen cyanide	2.93 D.
Hydrogen bromide	0.78	Water	1.84	Sulfur dioxide	1.6
Hydrogen iodide	0.38	Hydrogen sulfide	1.10	Hydrogen peroxide	2.1
ALIPHATIC COMPOUNDS.					
Alkyl chlorides	2.05	Alcohols	1.7	Nitriles	3.6
Alkyl bromides	2.05	Ethers	1.15	Nitro-compounds	3.4
Alkyl iodides	1.9	Ketones	2.7	Primary amines	1.3
AROMATIC COMPOUNDS					
Chlorobenzene	1.73	Phenol	1.7	Benzonitrile	4.37
Bromobenzene	1.71	Diphenylether	1.15	Nitrobenzene	4.23
Iodobenzene	1.2	Benzophenone	2.95	Aniline	1.56

expressed in debye units (10^{-18} e.s.-cm.). In addition to the data given in the table, it may be noted that hydrogen, nitrogen, carbon dioxide, carbon disulfide, iodine, bromine, chlorine, stannic chloride and iodide, boron trichloride, carbon tetrachloride, paraffins, cyclohexane and benzene have zero moments. In fact all molecules with a center of symmetry have no resultant dipole moment, as is to be expected from general considerations. It follows, therefore, that whereas carbon dioxide must form a linear molecule, sulfur dioxide and water, both of which have relatively

* The energy of dipole association is relatively small in comparison with the thermal energy of the molecules, and so the extent of this type of association is probably small except at very low temperatures.

large moments, must have angular structures. Further, since boron trichloride is nonpolar, the molecule is probably flat, so that it has a center of symmetry. On the other hand, phosphorus trichloride, and analogous halides of elements of the fifth group, as well as the hydrides, e.g., ammonia, are polar, and hence the molecule is probably in the form of a triangular pyramid with the nitrogen or similar atom at the apex.⁹

Dipole Moments of Linkages.—Since the dipole moment arises because of the difference in the electronegativity of two atoms connected by a chemical bond (p. 108), it is evident that it should be possible to associate a dipole moment with every linkage. The resultant dipole moment of a molecule is then the *vector* sum of the individual bond moments, since each acts along the line connecting the centers of the atoms forming the bond (J. J. Thomson, 1923). This view is in harmony with the fact that symmetrical molecules like carbon tetrachloride have no moment, although each carbon-chlorine bond has an appreciable moment, probably about 1.6 D.; the four bond moments are directed tetrahedrally and their vector sum is zero. A. Eucken and L. Meyer (1930) analyzed the dipole moments of a number of compounds, and suggested the following moments for various linkages:

H—C	C—O	C—Cl	H—O	C=O
0.4	0.7	1.5	1.6	2.3 D.

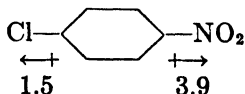
In each case the left-hand atom, as written, is less electronegative than the other, and so forms the positive end of the dipole. By the vector addition of these bond moments the resultant dipole moments of various substances were calculated and found to be in agreement with the observed values.

Group Moments.—For many purposes it is more convenient to consider the resultant moment of a group rather than those of the individual bonds. It will be seen from Table 88, for example, that one value of the dipole moment is associated with all aliphatic nitro-compounds, one with all nitriles, and so on; this implies that the determining factor is the polar group in each case. As a general rule the first member of each series has a somewhat lower moment, the subsequent increase being due to an induced moment in the hydrocarbon chain. Among aromatic compounds, too, each group contributes a definite moment, and the values quoted below were derived by J. W. Williams (1928), from measurements in solution, on the assumption that the C—H linkage in benzene has no dipole moment.

NO ₂	CN	OH	Cl	H	CH ₃	NH ₂
-3.9	-3.8	-1.7	-1.5	0	+0.4	+1.5 D.

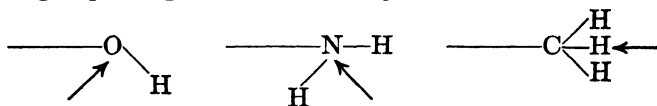
The signs preceding the group moments indicate the direction in which the moment operates. The nitro-group, for example, is known to be strongly electronegative, and so will presumably become negatively charged with respect to the benzene ring; the group moment is, therefore,

given a negative sign. The signs of the other groups may be determined (K. Højendahl, 1926) by substitution in the *para*-position to the nitro-group and measuring the resultant moment. The dipole moments of nitrobenzene and chlorobenzene, in benzene solution, are 3.9 and 1.5 d., respectively; that of *p*-chloronitrobenzene is 2.6 d., showing that the sign of the chlorine moment must be negative like that of the nitro-group. This may be seen more clearly with the aid of the diagram below; in accordance with the suggestion of N. V. Sidgwick (1930) the direction of the dipole is marked by an arrow with its point toward the negative end.



It is obvious that the resultant dipole should be about 2.4 d., in fair agreement with experiment. If the chlorine group moment had been positive, the *p*-chloronitrobenzene would have had a dipole moment of over 5 d. That the moments of the nitro- and methyl-groups have opposite signs is evident from the fact that *p*-nitrotoluene has a dipole moment of about 4.4 d. in solution.

In the pictorial representation above it has been assumed that the bond moments act along the line joining the chlorine or nitrogen atom to the carbon atom of the benzene ring. This can be justified in many ways, and the same can be said for the $-\text{CN}$ and $-\text{CH}_3$ groups, but not for $-\text{OH}$ and $-\text{NH}_2$. In these two cases the arrangement of the hydrogen atoms is not symmetrical about the line of attachment of the group to the benzene nucleus; the configurations of $-\text{OH}$ and $-\text{NH}_2$ groups, relative to the plane of the ring, may be represented as follows, that for the $-\text{CH}_3$ group being included for comparison.

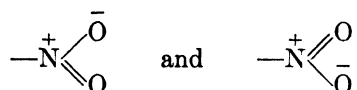


The direction of the resultant group moment is shown by an arrow in each case; for both $-\text{OH}$ and $-\text{NH}_2$ this makes an angle with the plane of the benzene nucleus. To illustrate the effect of this deviation, the dipole moment of *p*-chlorophenol may be considered. If the moments of the $-\text{Cl}$ and $-\text{OH}$ groups were collinear, as is the case for $-\text{Cl}$ with $-\text{CN}$, $-\text{NO}_2$ or $-\text{CH}_3$, the dipole moment of the compound would be about 0.2 d., i.e., 1.7 for $-\text{OH}$ minus 1.5 for $-\text{Cl}$, whereas the actual value is 2.4 d. By graphical construction, or by means of the equation given below, it can be shown that the vector representing the direction of the resultant moment of the $-\text{OH}$ group must consequently make an angle of 83° with that of the $-\text{Cl}$ moment.

When two identical groups are substituted in the *para*-positions, the moment of the compound should be zero, provided the groups are of the type not involving the direct attachment of oxygen or nitrogen to the

benzene ring; this has been confirmed for a number of substances, e.g., *p*-dichloro-, *p*-dibromo-, *p*-dinitro- and *p*-dimethylbenzene. For derivatives of the type *p*-ROC₆H₄OR and *p*-R₂NC₆H₄NR₂, however, the dipole moments are not zero; this result provides further support for the view that the moments of -OR and -NR₂ groups do not act in the direction of the C₆H₅-O or C₆H₅-N bond, respectively.

It may appear surprising that the -NO₂ group does not behave in the same manner as the -NH₂ and similar groups. From the fact that *p*-dinitrobenzene is effectively a nonpolar molecule it is clear that the moment of the -NO₂ groups acts in the plane, and along the symmetry axis, of the benzene ring. The explanation is that two configurations, viz.,



are possible, so that resonance occurs. The actual structure is thus a combination of the two forms and must consequently be symmetrical. The group moment will, therefore, act in the direction of the fourth valence bond of the nitrogen atom, and the nitro-group behaves like -CH₃, -Cl or similar groups. It will be seen, incidentally, that the negative nature of the -NO₂ group is really due to the N-O bonds.

Vector Addition of Moments.—The dipole moment of a molecule may be evaluated by adding vectorially the various group or bond moments; this can be done graphically, or by means of the equation

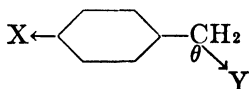
$$\mu^2 = \mu_1^2 + \mu_2^2 + 2\mu_1\mu_2 \cos \theta, \quad (60)$$

where μ is the resultant of two group or bond moments μ_1 and μ_2 , the values including the appropriate sign, which make an angle θ with each other. A simple application of (60) is in connection with the moments of di-substituted benzene derivatives. Since the benzene ring is flat, the angle between the dipole vectors of two *ortho*-groups is 60°, for two groups in the *meta*-position it is 120°, and for two *para*-groups the angle is 180°, provided the groups are not of the -OR or -NR₂ type, where R is hydrogen or an alkyl or similar radical. As a general rule, the agreement between the observed dipole moments and those calculated by means of (60), utilizing the group moments given above, is satisfactory, but for *ortho*-derivatives considerable discrepancies have been observed, e.g., for *o*-dinitrobenzene the values are 6.05 D. (observed) and 6.75 D. (calculated). The so-called **ortho-effect** is due partly to repulsion of the adjacent groups, leading to an increase in the angle between the dipole vectors, but the chief cause of the discrepancy is that each dipolar group induces a moment, in the opposite direction, in the other group; the observed resultant moment is thus considerably less than would have otherwise been the case (H. M. Smallwood and K. F. Herzfeld, 1930). Another instance of dipole induction is provided by chloroform: if the four atoms are ar-

ranged at the corners of a regular tetrahedron about the carbon atom, chloroform should have the same dipole moment as methyl chloride,* whereas the actual values are 1.1 and 1.85 D., respectively. It was thought, at one time, that the repulsion of the chlorine atoms resulted in a widening of the angles between the carbon-chlorine bonds, but this is known to be incorrect (p. 595); the low dipole moment of chloroform is undoubtedly due to the moment, acting in the opposite sense, induced by each C—Cl dipole in the other two (C. P. Smyth, 1933).¹⁰

Molecular Structure.—A few illustrations of the information concerning the configurations of simple molecules derived from dipole data were given on p. 551; another obvious conclusion is the planar hexagonal structure of the benzene ring, as shown by the agreement between observed and calculated moments of *m*- and *p*-derivatives. The zero moment of symmetrical *pp'*-biphenyl compounds supports the accepted linear structure of biphenyl. Dipole moment measurements permit of an immediate distinction between *cis*- and *trans*-dihalogeno-ethylene compounds; the *trans*-derivatives have a center of symmetry and consequently are nonpolar, whereas the *cis*-forms have relatively large moments (J. Errera, 1926). It is possible, in a similar manner, to distinguish between *syn*- and *anti*-oximes; the α -oxime of *p*-nitrobenzophenone has a large moment, while that of the β -form is small, and so it is evident that in the former the $^+N-O^-$ dipole of the oxime grouping is operating in the same direction as that in the nitro-group, whereas in the latter the reverse is true. The α -oxime must, therefore, be the *syn*-form and the β - is the *anti*-form.

Useful information concerning bond angles can be derived from dipole moments; this is, of course, a consequence of the vector additivity of bond and group moments. The bond angle of carbon may be calculated from measurements on *p*-benzyl compounds of the type



where X may be $-NO_2$, $-Cl$ or $-Br$, and Y can be $-Cl$, $-Br$ or $-CN$. If μ_1 is the group moment of X, and μ_2 that of $-CH_2Y$, and the resultant moment of the compound is μ , then it follows from (60) that

$$\cos \theta = (\mu^2 - \mu_1^2 - \mu_2^2)/2\mu_1\mu_2, \quad (61)$$

where θ is the bond angle of carbon. In this deduction it is assumed that the $-CH_2Y$ moment acts along the line of the C—Y bond, as shown;

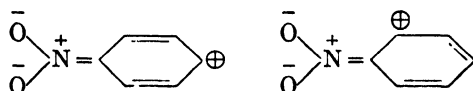
* It can be readily shown, and it is in fact obvious from a consideration of a symmetrical molecule such as carbon tetrachloride, that three identical groups at the corners of a regular tetrahedron are equivalent to one group of the same kind acting along the line joining the fourth corner to the central atom. The three C—Cl bond moments in $CHCl_3$ are thus equivalent to the one in CH_3Cl , whereas the three C—H moments in the latter are equivalent to the one in the former compound.

this can be justified, provided the value of μ_2 is taken to be that for the compound $C_6H_5CH_2Y$. The dipole moment (μ) of *p*-nitrobenzyl chloride ($NO_2C_6H_4CH_2Cl$) in benzene solution is 3.59 D., while the group moment of $-NO_2$, i.e., μ_1 , is 3.9 and the moment of benzyl chloride, i.e., μ_2 , is 1.85 D.; substitution of these values in (61) gives $\theta = 113.5^\circ$, in fair agreement with the tetrahedral angle (C. P. Smyth and W. S. Walls, 1932). The bond angles of oxygen and sulfur have been determined in an analogous manner, utilizing the dipole moments of *p*- or *pp'*-derivatives of diphenyl ether and thioether, respectively; in these cases an allowance must be made for the transmission of polarization effects through the nucleus by the substituent groups.¹¹

Dipole Moments and Resonance.—In a number of instances the observed dipole moment differs appreciably from the value expected from the individual group moments, and there is reason for believing that ionic structures, whose moments are quite different from that of the normal molecule, contribute to the actual configuration by resonance. An interesting aspect of this matter is presented by certain simple benzene derivatives. It was pointed out by L. E. Sutton (1931) that *ortho*- and *para*-directing groups, e.g., halogens, have smaller moments when substituted in a benzene ring than when attached to a saturated aliphatic group, although the reverse is the case for *meta*-directing groups, e.g., $-NO_2$ and $-CN$; that this is so may be readily seen from the data in Table 88. It appears, therefore, that the phenyl group tends to draw electrons away from the halogen atoms, thus decreasing their effective negative moment, whereas electrons are repelled by the nucleus toward the $-NO_2$ or $-CN$ group; this is in agreement with the permanent part of the "electromeric" effect, also called the "mesomeric" effect, postulated by organic chemists to account for the influence of various substituents on the reactivity of organic compounds. It is now widely accepted that the mesomeric effect is due to resonance between the normal molecule and such structures as



when X is a halogen. On the other hand when the substituent is a nitro-group the configurations



are possible in addition to the normal structure (C. W. Porter, 1934). An examination of these formulae shows that in the halogen compound the ionic structures contribute dipoles acting in the opposite sense to that of the normal halogen, while the reverse is true for the nitro-derivative. Further, in the former case resonance results in the availability of electrons in the *ortho*- and *para*-positions, but in the latter case there is a deficiency of electrons in these positions. The assumption of resonance, therefore, accounts both for the dipole moment effects and the chemical properties of the two groups of substituents.¹²

THE KERR EFFECT

The Kerr Constant.—The phenomenon of electrical double refraction, discovered by J. Kerr (1875) and known as the **electro-optical Kerr effect**,* has become of interest to chemists in recent years, although its applications lie mainly in the realm of physics. If a transparent isotropic substance, e.g., a liquid or gas, is placed in a stationary electric field, it becomes optically anisotropic and doubly refracting (cf. p. 346). In its simplest terms the theory of the Kerr effect is as follows. The electric field tends to bring about a definite orientation of the molecules, partly because of the moments induced in them and partly because of the permanent moments they may possess. Although the orienting tendency is opposed, to some extent, by the thermal motion of the molecules, there will be nevertheless a more or less regular arrangement under the influence of the applied field. If the molecules are not perfectly symmetrical, it is clear that a difference in the velocity of light, and hence in the refractive indices, in two directions at right angles is not unexpected. If n is the refractive index of the incident light of wave length λ , and n_p and n_s are the values for the emergent polarized light whose electric vectors vibrate parallel and perpendicular to the direction of the field † of strength E , then

$$B = \frac{n_p - n_s}{n\lambda} \cdot \frac{1}{E^2}, \quad (62)$$

where B is a constant. In general B is independent of wave length so that λ may be left out, and (62) reduces to

$$K = \frac{n_p - n_s}{n} \cdot \frac{1}{E^2}, \quad (63)$$

K being called the **Kerr constant** for the given substance.

When a molecule is polarizable to varying degrees in different directions, the polarizabilities along three imaginary principal axes at right angles in the molecule are b_1 , b_2 and b_3 , the mean molar polarizability of the molecule α being $\frac{1}{3}(b_1 + b_2 + b_3)$. As a first approximation the molecule may be regarded as an optical "polarization ellipsoid," with its semi-axes in the ratio of $b_1 : b_2 : b_3$. By a theoretical treatment, having features in common with that described for dipole moments (p. 544), it has been deduced, assuming a random distribution of completely non-associated molecules, e.g., gas or vapor, that

$$K = \frac{(n^2 - 1)(n^2 + 2)}{4n^2} \cdot \frac{\Theta_1 + \Theta_2}{\alpha} \left(\frac{D + 2}{3} \right)^2, \quad (64)$$

where D is the dielectric constant of the medium. The quantities Θ_1 and Θ_2 are defined as follows:

$$\Theta_1 = \frac{1}{45kT} [(a_1 - a_2)(b_1 - b_2) + (a_2 - a_3)(b_2 - b_3) + (a_3 - a_1)(b_3 - b_1)] \quad (65)$$

* The adjective is used to distinguish this from another effect discovered by Kerr.

† The electric vector vibrates at right angles to the plane of polarization of the light.

and

$$\Theta_2 = \frac{1}{45k^2T^2} [(\mu_1^2 - \mu_2^2)(b_1 - b_2) + (\mu_2^2 - \mu_3^2)(b_2 - b_3) + (\mu_3^2 - \mu_1^2)(b_3 - b_1)], \quad (66)$$

in which a_1 , a_2 and a_3 are the moments induced in the directions of the three axes by a stationary electric field, and μ_1 , μ_2 and μ_3 are the components of the permanent dipole moment μ in the same respective directions. It will be recalled (p. 534) that the induced moment is proportional to the polarizability, so that a_1/b_1 , a_2/b_2 and a_3/b_3 are equal. Since (64) is strictly applicable to gases and vapors, both the refractive index n and the dielectric constant D differ little from unity; further, introducing the value for the mean polarizability α in terms of the molar refraction [equation (26)], the equation for K becomes

$$K = \frac{3\pi N\rho}{M} (\Theta_1 + \Theta_2) = K_1 + K_2. \quad (67)$$

The Kerr constant can thus be divided into two terms; the **anisotropy term** K_1 is determined by Θ_1 , and the **dipole term** K_2 is proportional to Θ_2 . It is seen from (65) and (66) that K_1 varies inversely with the temperature, whereas K_2 is proportional to $1/T^2$. For a nonpolar molecule Θ_2 is zero, since μ is zero, so that the Kerr constant K is inversely proportional to the absolute temperature.

The Anisotropy Term.—From consideration of the scattering of light, R. Gans (1921) deduced that

$$\frac{a}{b} = \frac{n_\infty^2 - 1}{n^2 - 1} \approx \frac{n_\infty - 1}{n - 1}, \quad (68)$$

where n_∞ is the refractive index for infinite wave length, the infrared contribution being included. Since, as seen above, a/b has the same value for the three axes, (65) may be simplified to

$$\Theta_1 = \frac{1}{45kT} \cdot \frac{n_\infty - 1}{n - 1} [(b_1 - b_2)^2 + (b_2 - b_3)^2 + (b_3 - b_1)^2]. \quad (69)$$

After scattering by any medium light is found to be polarized; if I_p is the intensity of the ray with vibrations parallel to the incident light, and I_s with vibrations perpendicular to that direction, then the ratio is called the **depolarization factor**, Δ , i.e.,

$$\Delta = I_p/I_s. \quad (70)$$

The theory of light scattering leads to an equation relating Δ to the polarizabilities along the three principal axes; thus,

$$\frac{10\Delta}{6 - 7\Delta} = \frac{(b_1 - b_2)^2 + (b_2 - b_3)^2 + (b_3 - b_1)^2}{(b_1 + b_2 + b_3)^2}. \quad (71)$$

Remembering that $b_1 + b_2 + b_3$ is equal to 3α , combination of this result with (69) gives

$$\Theta_1 = \frac{1}{2kT} \cdot \frac{(n_\infty - 1)(n - 1)}{(\pi N\rho/M)^2} \cdot \frac{\Delta}{6 - 7\Delta}, \quad (72)$$

the approximations for α , valid for gases [cf. equation (37)], being introduced. By means of (72) it is thus possible to evaluate the anisotropy portion K_1 of

the Kerr constant from measurements of light scattering whereby Δ is determined. It may be noted that since the refractive index is in general greater than unity, the right-hand side of (69) must be positive, so that K_1 has always a positive sign.

The Dipole Term.—If the total Kerr constant is obtained experimentally subtraction of K_1 gives the dipole contribution; although the anisotropy term cannot be determined with accuracy, it is generally so much smaller than the dipole term that the error is of no significance. The quantity Θ_2 is related to the three principal polarizabilities and to the components of the dipole moment by (66); this is complicated and in a few cases some simplification is possible. For example, if the resultant dipole of the molecule lies in the plane made by the axes designated by the suffixes 1 and 3, and makes an angle ϕ with the latter, then μ_2 is zero, and

$$\mu_1 = \mu \sin \phi \quad \text{and} \quad \mu_3 = \mu \cos \phi, \quad (73)$$

and (66) becomes

$$\Theta_2 = \frac{1}{45k^2T^2} \mu^2 [3 \cos^2 \phi (b_3 - b_1) + 2b_1 - b_2 - b_3]. \quad (74)$$

In the special case when the moment is parallel to the direction of the axis 3, i.e., $\phi = 0$, then

$$\mu_3 = \mu \quad \text{and} \quad \mu_1 = \mu_2 = 0;$$

under these conditions

$$\Theta_2 = \frac{1}{45k^2T^2} \mu^2 (2b_3 - 2b_2 - b_1). \quad (75)$$

Sign of Kerr Constant.—As seen above, K_1 must be positive and so if the Kerr constant is negative it follows that the dipole term is negative; the circumstances under which this can arise may be deduced from the following considerations. Suppose the molecule has an axis of symmetry, so that as far as polarizability is concerned it is an ellipsoid of rotation; the polarizabilities will then be the same along two axes, e.g., $b_1 = b_2$, and hence $a_1 = a_2$. If the axes are chosen so that the dipole is perpendicular to axis 2 and makes an angle ϕ with axis 3, then as before μ_2 is zero, and

$$\mu_1 = \mu \sin \phi \quad \text{and} \quad \mu_3 = \mu \cos \phi, \quad (76)$$

so that equation (74) is applicable, but as b_1 and b_2 are now equal, it is seen that

$$\Theta_2 = \frac{1}{45k^2T^2} \mu^2 [(3 \cos^2 \phi - 1)(b_3 - b_1)]. \quad (77)$$

Suppose $\phi = 0$, so that the dipole acts in a direction parallel to the 3-axis, as in Fig. 119, I; then $\cos \phi$ is unity, and

$$\Theta_2 = \frac{1}{45k^2T^2} 2\mu^2(b_3 - b_1). \quad (78)$$

If axis 3 is that of maximum polarizability, as indicated in the diagram, then $b_3 > b_1$, and Θ_2 is positive. If, on the other hand, ϕ is 90° , as in Fig. 119, II, then $\cos \phi$ is zero, and

$$\Theta_2 = -\frac{1}{45k^2T^2} \mu^2(b_3 - b_1), \quad (79)$$

so that b_3 being greater than b_1 the dipole term of the Kerr constant must be negative. The value of Θ_2 changes from positive to negative when $3 \cos^2 \phi$ is equal to unity, i.e., when ϕ is 54.7° ; for larger values of ϕ the term Θ_2 , and hence the Kerr constant, becomes increasingly negative. It can be stated, therefore, that when the resultant dipole of a molecule is at right angles to the axis of maximum polarizability, or close to it, the Kerr constant of the substance will be negative.

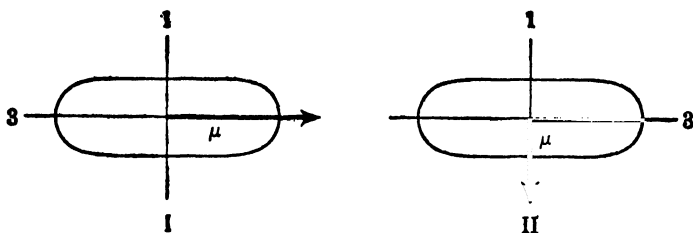


FIG. 119. Direction of dipole and sign of Kerr constant

The Polarization Ellipsoid.—The relative “dimensions” of the polarization ellipsoid can be evaluated if three equations are available relating b_1 , b_2 and b_3 to experimentally determinable quantities. One of these is (71), involving the depolarization factor, and another is (75), the axes being chosen so that the dipole lies along one of them. The third relationship is supplied by the molar refraction, since $[R]$ is equal to $4\pi N\alpha/3$, and α is equal to $\frac{1}{3}(b_1 + b_2 + b_3)$. By combining these equations, H. A. Stuart (1929 *et seq.*) has been able to calculate the polarizabilities of a number of molecules in three directions at right angles, one corresponding to that of the resultant dipole. For substances consisting of oxygen, carbon and nitrogen, which have atomic refractions, and hence polarizabilities, of the same order, together with hydrogen, the shape of the polarization ellipsoid will approximate to the shape of the molecule. A knowledge of the quantities b_1 , b_2 and b_3 can therefore be employed to provide information concerning molecular structure.

Experimental Results.—In the determination of the Kerr constant the incident beam is polarized in a plane making an angle of 45° with the direction of the electrical field; when passing through a cell containing the experimental substance subjected to a high voltage, the light becomes circularly polarized as a result of the Kerr effect. The phase difference between the components perpendicular and parallel to the lines of force is proportional to the Kerr constant. By measuring the intensity of the emergent light, with a Nicol prism as analyzer at different angles relative to the polarizer, the phase difference and hence the Kerr constant K can be evaluated. The anisotropy term K_1 is obtained, as shown above, from observations on the scattering of light. Some results for a number of gases are recorded in Table 89.

Applications.—The Kerr constant has been used mainly to confirm data obtained from other sources, although it has provided direct information about molecular structure in certain instances where free

TABLE 89. KERR CONSTANTS OF GASES

Substance	$K \times 10^{15}$	$K_1 \times 10^{15}$	$K_2 \times 10^{15}$	μ
CO ₂	1.42	1.42	0	0 D.
CS ₂	21.0	21.0	0	0
HCl	5.75	0.44	5.31	1.03
SO ₂	-9.2	1.75	-10.95	1.6
CHCl ₃	-7.5	1.1	-8.6	1.05
CH ₃ Cl	36.5	0.8	35.7	1.86
(C ₂ H ₅) ₂ O	-3.9	3.7	-7.6	1.2

rotation is possible. The negative value for sulfur dioxide, for example, shows that the molecule cannot be linear; if it had an unsymmetrical linear configuration, and so possessed a dipole moment, the direction of the latter would have to be parallel to the longer axis of the molecule and hence to that of maximum polarizability. The small Kerr constants of carbon dioxide and nitrous oxide, consisting entirely of the anisotropy term, confirm the linear structure of these substances. The surprisingly high value for carbon disulfide may be explained by the large polarizability of the bond between carbon and sulfur. The results for the nonpolar hydrocarbons and carbon tetrachloride, which have very small Kerr constants, and for the polar compounds ammonia and hydrogen chloride, are such as might have been anticipated.

Among the compounds whose Kerr constants have been studied in the vapor state, those with negative values are of special interest. A comparison of the structures of methyl chloride and chloroform accounts for the difference in sign. Since the carbon and chlorine atoms are the most polarizable parts of the molecule (see Table 82), the direction of maximum polarizability (b_3) in methyl chloride is the same as that of the resultant dipole moment (μ), as shown in Fig. 120, A, whereas in

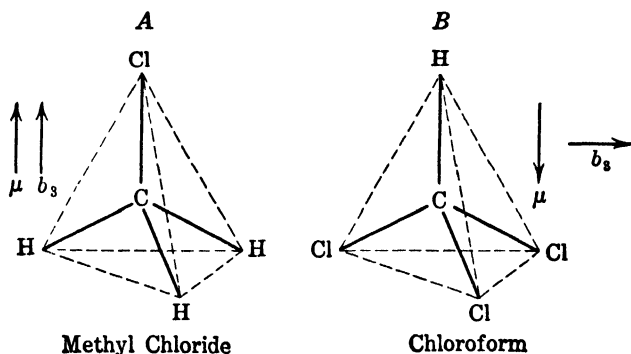


FIG. 120. Positive and negative Kerr constants

chloroform the two directions are at right angles (Fig. 120, B). In accordance with the generalization derived above, a negative Kerr constant for chloroform is to be expected.

In diethyl ether there is a possibility of rotation about carbon-oxygen bonds, and the two extreme positions are represented by I and II in Fig. 121. Since the Kerr constant is negative, it is clear that position I is favored, for the direction of maximum polarizability for such a configuration would be at right angles to that of the resultant dipole moment. The polarizabilities b_1 , b_2 and b_3 , the axis of the latter being perpendicular to the moment, were found to be 113 , 71 and 79×10^{-25} respectively; since the polarizability is much greater in one direction than in the other two, the molecule must be elongated, as represented by structure I.

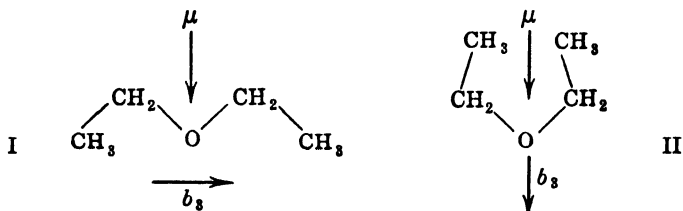


FIG. 121. Configuration of diethyl ether

In the series of homologous ketones, the Kerr constant, which is positive for acetone, decreases, and di-isopropyl ketone has a negative value, as the following results show:

$K \times 10^{15}$	CH_3COCH_3 32.1	$\text{CH}_3\text{COC}_2\text{H}_5$ 16.0	$\text{C}_2\text{H}_5\text{COC}_2\text{H}_5$ 5.2	$\beta\text{-C}_3\text{H}_7\text{COC}_3\text{H}_7$ -15.3
--------------------	------------------------------------	---	---	---

In acetone the maximum polarizability must be in the direction of the $\text{C}=\text{O}$ bond, and this is also that of the dipole moment; the Kerr constant is, therefore, positive. In isopropyl ketone, however, the accumulation of carbon atoms increases the polarizability perpendicular to the carbonyl group, and a negative Kerr constant results.

Liquids.—The Kerr constants of liquids are known only approximately, and the anisotropy term cannot be determined with any certainty; it is the general practice to give the value of the constant K relative to that of carbon disulfide, as in Table 90.

TABLE 90. KERR CONSTANTS OF LIQUIDS RELATIVE TO CARBON DISULFIDE

Substance	K/K_{CS_2}	μ	Substance	K/K_{CS_2}	μ
CCl_4	0.023	0 D.	$\text{C}_6\text{H}_5\text{Cl}$	3.85	1.7 D.
C_6H_6	0.12	0	$\text{C}_6\text{H}_5\text{NO}_2$	60.0	4.2
$\text{C}_2\text{H}_5\text{Cl}$	3.00	1.9	$\text{C}_6\text{H}_5\text{NH}_2$	-0.38	1.5
CH_3NO_2	3.30	3.4	$p\text{-CH}_3\text{C}_6\text{H}_4\text{NH}_2$	-1.28	1.9

Applications.—The equations deduced above apply only to gases, and possibly to dilute solutions, but the general conclusion concerning the sign of the Kerr constant and the direction of the dipole in the molecule appears to be applicable to pure liquids. Aniline and the toluidines give

negative values, and since the polarizability of these molecules is greatest in the plane of the benzene ring, the resultant dipole must act almost at right angles (Fig. 122, A); this is in harmony with the accepted view that in an amine the three valences of the nitrogen atom do not lie in one plane (cf. p. 553). In the molecule of nitrobenzene, however, the direction of the dipole coincides with the direction of maximum polarizability (cf. p. 554) and the Kerr constant is consequently positive (Fig. 122, B). The resultant moments of the phenolic ethers make an

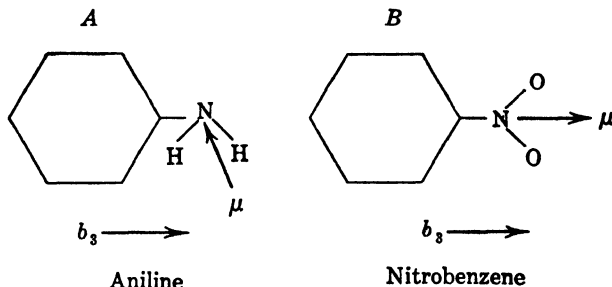


FIG. 122. Kerr constants of aniline (negative) and nitrobenzene (positive)

angle with the plane of the benzene ring; negative or small positive Kerr constants are to be expected. The observed values are positive, but smaller than is usual for substances with dipole moments of the same order of magnitude.¹³

MOLECULAR SPECTRA OF GASES

Electronic Spectra.—The spectra of molecules are often called **band spectra** to distinguish them from the line spectra of atoms. It is true that molecular spectra of gases appear in the form of bands, but further analysis shows them to consist of a large number of closely spaced lines. In order to account for the structure of band spectra it has been necessary to assume that the electronic, vibrational and rotational energies of molecules are quantized (cf. p. 31). A molecule, like an atom, can exist in a number of electronic states or energy levels, and the change from one particular state to another results in the absorption or emission of a definite amount of energy, and a line of definite frequency is produced in the spectrum. The critical potential of a molecule, which is a measure of the energy difference of two electronic states, is of the order of 5 e.v., i.e., about 8×10^{-12} erg per molecule. Insertion of this value in the quantum theory equation $\epsilon = h\nu$, where h is 6.6×10^{-27} erg-sec., shows that the frequency of the radiation accompanying an electronic transition is about 1.2×10^{15} vibrations per sec., which is equivalent to a wave length of 2500 Å. The so-called **electronic spectra** should thus appear in the short wave region of the visible or in the near ultraviolet part of the spectrum.

In each electronic state the molecule can have a number of vibrational sublevels, the energy difference between them being about one-tenth to one-hundredth that between two successive electronic levels. Further, each vibrational level has a set of rotational sublevels with energy separations of the order of one-hundredth that between two vibrational levels. The various levels may be represented by the diagram in Fig. 123; *A* and *B* are two electronic levels, each having a series of vibrational

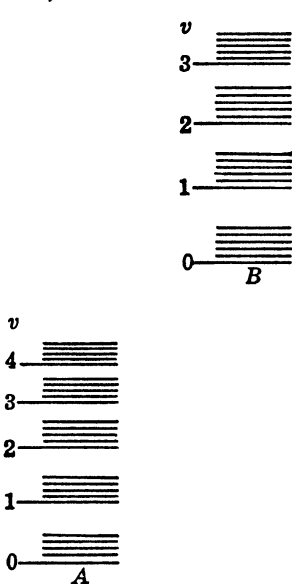


FIG. 123. Energy levels

sublevels, indicated by $v = 0, 1, 2$, etc., and every one of the latter has a number of rotational levels, for which $J = 0, 1, 2$, etc.; v and J are the vibrational and rotational quantum numbers respectively. When a normal molecule in the state *A* receives energy and is thus converted into the excited state *B* then, provided there were no changes in the vibrational and rotational energies, one line would appear in the absorption spectrum. In practice, however, each electronic change is accompanied by a vibrational change, the magnitude of which is determined by the initial and final vibrational levels, and if the vibrational-energy quantum is about 0.1 e.v., compared with 5 e.v., for the electronic energy, the spectrum appearing in the region of 2500 Å. will consist of a number of lines with a spacing of approximately 50 Å. Simultaneously with electronic and vibrational energy changes, there will be rotational transitions; for each definite change in electronic and vibrational

energies, therefore, every rotational transition will produce a line in the spectrum. Since rotational energy changes are small, viz., of the order of 0.005 e.v., the lines will be closely spaced. The set of rotational lines corresponding to a particular vibrational change is called a band, and every vibrational transition gives a separate band of this type; the whole set of bands for a given electronic transition is called a **band group** or **band system**. Each electronic change occurring under given conditions will yield such a band system, and the whole series of band groups forms the electronic band spectrum of the gaseous molecule. It is clear that such spectra may be very complex, even if there is only one electronic transition. In general a nonlinear molecule can have $3n - 6$ different modes of vibration (see p. 331), where n is the number of atoms in the molecule, and in addition there will be rotation about three axes. The number of possible transitions is thus large, in spite of certain restrictions such as occur in atomic spectra. It is because of this complexity that diatomic molecules have been studied particularly, but progress is being made in the elucidation of the spectra of more complex substances.

Infrared Spectra.—If a change occurred in vibrational energy only, within one and the same electronic level, the energy difference, as seen above, would be about 0.1 e.v., i.e., 1.6×10^{-13} erg per molecule. According to the quantum theory the corresponding spectral line should have a frequency of approximately 2.5×10^{13} vibrations per sec., which would give a wave length of 120,000 Å. or 12 μ . The radiations accompanying purely vibrational changes should thus appear in the near infrared region of the spectrum. Actually the vibrational transition is always accompanied by rotational energy changes which give a series of closely spaced lines; **vibration-rotation** bands produced in this way are generally found in the wave length region of 1 to 23 μ . A diatomic molecule has one mode of vibration only, and so produces one set of such bands; for a polyatomic molecule, however, several sets are possible (p. 570).

The band spectra observed in the far infrared, at wave lengths of about 200 μ , result from energy changes of the order of 0.005 e.v.; these must be due to purely rotational transitions, the electronic and vibrational energies remaining unchanged. The **rotation spectra** obtained in this manner are obviously the simplest type of molecular spectra, but the far infrared region of the spectrum is unfortunately the one presenting the greatest experimental difficulties.

A molecule will interact with radiation provided the process is accompanied by a change of dipole moment; for a diatomic molecule this means that the substance must possess a resultant dipole moment if it is to exhibit spectra in the infrared. Consequently, homopolar diatomic molecules, e.g., H_2 , N_2 , O_2 , Cl_2 , etc., do not have pure rotation or vibration-rotation spectra. It may be noted, however, that an electronic transition is almost invariably accompanied by changes in the vibrational and rotational energies, even if the molecule is homopolar in its ground state. For polyatomic molecules some vibrations are "active," even if the substance has no resultant dipole moment (p. 570).

Experimental Methods.—For the study of electronic spectra the most convenient apparatus is the quartz spectrograph with a suitable photographic attachment. The emission spectrum can sometimes be excited by an electrical discharge through the gas or by a suitable chemical reaction, otherwise the absorption spectrum is examined. The source of light may then be a discharge tube containing hydrogen; under suitable conditions it can be made to give a practically continuous emission from 5000 Å. into the ultraviolet. Infrared spectra are invariably studied in absorption; the source of radiation is often a Nernst glower, consisting of a filament of cerium, thorium and zirconium oxides heated electrically to about 2000° c., for wave lengths up to about 10–15 μ . A commercial electric-heater unit, known as a "Glo-bar," consisting of a rod of silicon carbide, is coming into common use as it is more robust than the Nernst filament. For longer wave lengths the Welsbach gas mantle is generally employed. Spectrometers containing prisms of fluorite, rock-salt or sylvine may be used in the near infrared, viz., 3 to 9 μ , 8 to 16 μ , and 15 to 20 μ , respectively, but it appears that better results are obtained with diffraction gratings, e.g., of the "echelette" type. By means of specially treated plates the spectral region up to 1.3 μ can be studied photographically, but the vacuum thermocouple may be employed in this vicinity, or in any other part of the

infrared spectrum, to determine the intensities of the absorption regions: a form of the Boys radiometer is preferred by some workers because of its high sensitivity.¹⁴

Rotational Spectra.—It can be deduced from the Schrödinger equation (p. 75) that the rotational energy ϵ_J , of a diatomic molecule, around an axis perpendicular to the line joining the nuclei, is given by

$$\epsilon_J = \frac{h^2}{8\pi^2 I} J(J+1), \quad (80)$$

where I is the moment of inertia of the molecule about the axis of rotation, and J is the rotational quantum number which may be zero or integral. In the transition of the molecule from an upper level J' to a lower level J'' , the energy emitted is

$$\epsilon_{J'} - \epsilon_{J''} = \frac{h^2}{8\pi^2 I} [J'(J'+1) - J''(J''+1)], \quad (81)$$

assuming the moment of inertia to be the same in both levels. As a general rule rotational transitions are restricted to those in which $\Delta J = 1$, so that $J' - J'' = 1$ and (81) becomes

$$\epsilon_{J'} - \epsilon_{J''} = \frac{h^2}{8\pi^2 I} 2J'. \quad (82)$$

According to the quantum theory this is equivalent to $hc\bar{\nu}$, where $\bar{\nu}$ is the frequency, expressed in wave numbers (p. 37), of the radiation accompanying the rotational transition, and c is the velocity of light; hence,

$$\bar{\nu} = \frac{h}{4\pi^2 Ic} m = 2Bm, \quad (83)$$

where J' has been replaced by the symbol m , representing any integer, and B is used for the constant quantity $h/8\pi^2 Ic$. Owing to the variation of the moment of inertia, it is often necessary to include an additional term, viz., $-Dm^3$, where D is a small constant; for present purposes this correction may be neglected.

By putting m equal to 1, 2, 3, etc., it is seen from (83) that the frequencies of successive lines in the rotational band are $2B$, $4B$, $6B$, etc., and from a study of the spectrum it should be possible to estimate the value of B and hence the moment of inertia (I) of the molecule. For a diatomic molecule of the type under consideration the latter is equal to μr^2 , where μ^* is the "reduced mass" of the molecule (see p. 148) and r is the distance apart of the nuclei. The value of μ can be readily ascertained from the masses m_a and m_b of the constituent atoms, since $\mu = (m_a m_b)/(m_a + m_b)$, and so the distance between the atoms can be

* This should not be confused with the dipole moment; it is unfortunate that the same symbol is used.

calculated from measurements of the rotational spectra. The results can be obtained with greater accuracy from other types of spectra and so their consideration will be deferred. It should be noted that, since the masses of the atoms determine the separation of the rotational lines, an "isotope effect" is to be expected; the shift in the lines is, however, too small to be detected in the far infrared.

Vibrational-Rotational Spectra.—According to the equations of wave mechanics, the energy ϵ_v of a harmonic oscillator, that is, one in which the restoring force is proportional to the displacement (Hooke's law), is

$$\epsilon_v = (v + \frac{1}{2})h c \omega_e, \quad (84)$$

where v is the vibrational quantum number, which may be zero or an integer, and ω_e cm.⁻¹ is the equilibrium frequency of vibration of the oscillator, i.e., for small displacements. As a general rule the movement is not strictly harmonic, and so it is necessary to include at least one additional term; thus,

$$\epsilon_v = (v + \frac{1}{2})h c \omega_e - (v + \frac{1}{2})^2 h c x \omega_e, \quad (85)$$

where x is the "anharmonicity constant."

Theoretically, vibrational transitions, for a harmonic oscillator, are restricted to changes of unity in the quantum number. For anharmonic oscillations, however, the selection rules are such as to make other transitions possible, but with a smaller degree of probability. When the vibrational energy of a molecule changes from the initial level in which v is 0 to that in which v is 1, in absorption, the band produced is called the **fundamental**. The energy change may be readily derived from (85) as $\epsilon_1 - \epsilon_0 = (1 - 2x)h c \omega_e$, so that the frequency of the corresponding spectral line $\bar{\nu}$, which is the origin of the fundamental band, is given by

$$\bar{\nu} = (1 - 2x)\omega_e. \quad (86)$$

When the initial value of v is 0 and the final value is 2 in a transition, the energy change is $(1 - 3x)2h c \omega_e$, and the frequency of the spectral line, the origin of the **first overtone** band, is then $(1 - 3x)2\omega_e$. Similarly, the origin of the **second overtone*** band, produced when v changes from 0 to 3, or the reverse, is $(1 - 4x)3\omega_e$; changes in the vibrational quantum number greater than three are rare and need not be considered. It will be seen that the frequencies of the origins of the fundamental, and first and second overtone bands are approximately in the ratio of 1 to 2 to 3, so that the wave lengths are in the proportion of 1 to $\frac{1}{2}$ to $\frac{1}{3}$. This fact is important, as the overtone bands appear in regions of shorter wave length where they can be studied more readily. For example, hydrogen chloride has a fundamental at 3.46μ , and first and second overtones at 1.76μ and 1.19μ respectively; the band at 1.76μ has been subjected to very detailed examination.† If the positions of the origins of the

* The first, second, etc., overtones are sometimes referred to as the second, third, etc., harmonic bands, respectively.

† In this work a compromise must be struck, because the intensity of the band falls off in the overtones, the second overtone being often too weak to be studied.

bands are measured carefully, it is clearly possible to evaluate both ω_e and $x\omega_e$. For hydrogen chloride, for example, the frequencies, in wave numbers, of the centers of the three bands are 2885.9, 5668.0 and 8347.0 cm^{-1} ; from these it follows that ω_e is 2988.9 and $x\omega_e$ is 51.65 cm^{-1} .

For most purposes the anharmonicity of the oscillations may be neglected, and if it may be assumed that the rotational and vibrational energies of a diatomic molecule can be added, there being no interaction between the two kinds of motion, then from (80) and (84),

$$\epsilon = (v + \frac{1}{2})h\omega_e + \frac{h^2}{8\pi^2 I} J(J+1). \quad (87)$$

For a simultaneous vibrational and rotational transition from v' to v'' and from J' to J'' , it follows that

$$\epsilon' - \epsilon'' = h\omega_e(v' - v'') + \frac{h^2}{8\pi^2 I} [J'(J'+1) - J''(J''+1)]. \quad (88)$$

For the fundamental band $v' - v''$ is unity, and the change in the rotational quantum number, as in pure rotation spectra, is restricted to unity. If $J' - J''$ is $+1$, then (88) becomes

$$\epsilon' - \epsilon'' = h\omega_e + \frac{h^2}{4\pi^2 I} J', \quad (89)$$

where J' can be 1, 2, 3, 4, etc., but not zero, for in the latter case J'' would be negative. On the other hand, if $J' - J''$ is -1 , then (88) takes the form

$$\epsilon' - \epsilon'' = h\omega_e - \frac{h^2}{4\pi^2 I} (J' + 1), \quad (90)$$

where J' can now be 0, 1, 2, etc. The two relationships may be combined in the form

$$\epsilon' - \epsilon'' = h\omega_e + \frac{h^2}{4\pi^2 I} m, \quad (91)$$

where m can be $\pm 1, \pm 2, \pm 3$, etc., but *not* zero. The frequency, in wave numbers, of the spectral line corresponding to the energy change is obtained on dividing by hc , so that

$$\bar{\nu} = \omega_e + \frac{h}{4\pi^2 I c} m = \omega_e + 2Bm, \quad (92)$$

where B has the same significance as on p. 566. The first term on the right-hand side gives the origin, or center, of the fundamental band, whereas the second term gives the rotational fine structure. If m is positive then a series of lines on the higher frequency, i.e., short wave, side of the origin is obtained; these constitute what is called the *R*-branch of the vibration-rotation band. When m is negative the lines of the

P-branch are obtained; these lie on the long wave side of the center of the band. Since m cannot be zero in the case under consideration, the actual origin of the band is marked by a gap. For molecules possessing angular momentum about the axis joining the nuclei, it is possible for m to be zero; that is to say, the vibrational energy change can occur unaccompanied by any rotational transition. Since the moment of inertia is appreciably different in the initial and final states, the second term in (88) is not zero, although J' and J'' are equal. There are thus very small changes in rotational energy which give rise to a series of very closely spaced lines, known as the *Q*-branch, which appears at the origin of the band. It is frequently observed for polyatomic molecules but not for diatomic molecules, with the exception of nitric oxide which has an odd electron.

It can be readily shown, by the method used above, that the frequencies of the rotational lines in any vibration-rotation band, neglecting anharmonicity, are given by

$$\bar{\nu} = n\omega_e + 2Bm, \quad (93)$$

where n is 1 for the fundamental, and 2 and 3 for the first and second overtones, respectively. It follows, therefore, that in any of these bands the frequency separation, in wave numbers, of the successive rotational lines is equal to $2B$, i.e., $h/4\pi^2 Ic$; hence it should be possible to calculate the moment of inertia and the equilibrium distance between the nuclei of a diatomic molecule from a study of band spectra. Owing to the fact that the moment of inertia does not remain unchanged the frequency separation is not constant, and so a mean value is generally employed.

For harmonic motion it is possible to write

$$2\pi\omega_e c = (f/\mu)^{1/2}, \quad (94)$$

where f is the restoring force per cm. displacement acting between the nuclei. If the equilibrium vibration frequency is known, either from (93) or, more accurately, after making an allowance for anharmonicity (p. 567), from measurements on the origins of the vibration-rotation bands, then the restoring force, or **force constant**, can be calculated. The results obtained from the spectra of the hydrogen halides are given in Table 91; the moments of inertia (I) and internuclear distances (r) are

TABLE 91. MOLECULAR CONSTANTS OF HYDROGEN HALIDES FROM SPECTRA

Substance	ω_e (cm. ⁻¹)	I (c.g.s.)	r (cm.)	f (dynes/cm.)
HF	4141	1.35×10^{-40}	0.92×10^{-8}	9.2×10^6
HCl	2989	2.65	1.28	4.8
HBr	2650	3.31	1.42	3.9
HI	2309	4.22	1.60	3.0

also included. The binding force between hydrogen and halogen decreases with increasing atomic weight of the latter; this accounts for the increasing readiness of the hydrogen halides to dissociate upon heating.

If the restoring force may be assumed to remain unchanged when an atom is replaced by its isotope, then it is evident from (94) that since μ is altered so also must be ω_e ; the positions of the lines in the vibration-rotation band will thus be shifted. The extent of the shift of the origin has been derived on p. 148, and the same value should apply to each rotational line. The splitting of the lines in the fundamental and first overtone bands of hydrogen chloride and in the fundamental of deuterium chloride, due to the isotopes ^{35}Cl and ^{37}Cl , has been observed; the isotopic shifts are in excellent agreement with the calculated figures.

Polyatomic Molecules.—A diatomic molecule has only one mode of vibration, and so it gives a comparatively simple spectrum. A molecule containing n atoms, where $n > 2$, has $3n - 6$ vibrational modes, of which $n - 1$ are stretching or valence vibrations, and $2n - 5$ are bending or deformation vibrations. For a linear molecule these values become $3n - 5$, $n - 1$ and $2n - 4$, respectively.* These vibrations are, however, not always different, for as the result of symmetry factors two or more vibrations may coincide; they are then said to be **degenerate**. A linear triatomic molecule should have four vibrational modes; one of these is doubly degenerate and

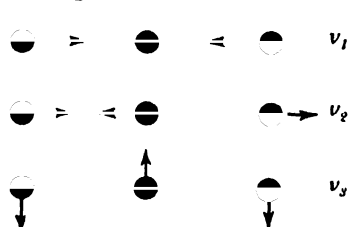


FIG. 124. Vibrational modes of a linear triatomic molecule

so there are only three different normal vibrations, as shown in Fig. 124. The corresponding frequencies are ν_1 , ν_2 and ν_3 , of which ν_3 is degenerate.† Of the other two vibrations ν_1 is said to be “symmetric” whereas ν_2 is “antisymmetric”: as a result of a symmetry operation the coordinates of the displacements remain unchanged in the former case, but are altered in sign in the latter. Another classification depends on the direction of change of the dipole moment

of the molecule accompanying the vibration. In this connection ν_1 and ν_2 are “parallel” vibrations and ν_3 is a “perpendicular” vibration, the changes being parallel and perpendicular, respectively, to the axis of symmetry of the molecule. For a symmetrical linear molecule, such as carbon dioxide, ν_1 is inactive in the infrared, because the electrical symmetry of the molecule remains unaltered during the vibration; that is to say, the dipole moment is not altered by the symmetrical displacements of the two outer atoms. For an unsymmetrical molecule, on the other hand, the value of the dipole moment does change in the course of the vibration, and ν_1 is active in addition to ν_2 and ν_3 . It must be pointed out that even if a particular oscillation does not appear as a fundamental band, it is generally evident in the form of a combination frequency with one of the other active vibrations.

A nonlinear triatomic molecule has three normal vibrations none of which is degenerate; these are shown in Fig. 125. The frequencies ν_1 and ν_2 are both symmetric and parallel, but ν_3 is antisymmetric and perpendicular. Since all three vibrations are accompanied by changes in the dipole moment of the

* The actual, complex, vibrations of a polyatomic molecule are treated as equivalent to $3n - 6$ (or $3n - 5$) normal vibrations, in each of which all the nuclei vibrate in phase and with the same frequency.

† The bending vibration occurs with the same frequency (ν_3) in two planes at right angles.

molecule, they are all active in infrared spectra. The more atoms in the molecule the greater the number of vibrational modes, and so the larger the number of vibration-rotation bands to appear in the spectrum. With water vapor, for example, seventeen bands, including fundamental, harmonic and combination bands, have been observed. It is obvious that the identification of the frequencies for even relatively simple polyatomic molecules is a matter of some difficulty. Nevertheless, as a result of experience and of the development of certain rules, expert workers in the field of spectroscopy have made remarkable progress in the study of relatively complex molecules, especially those having highly symmetrical configurations.

The rotational energy of any linear molecule is given by the same equation as for a diatomic molecule, since there is only one moment of inertia. For parallel vibrational frequencies, i.e., "parallel bands," the change ΔJ in the rotational quantum number is always ± 1 , but for perpendicular bands ΔJ may be zero, so that a Q -branch is possible. For nonlinear molecules there will be two or three different moments of inertia, and the rotational fine structure of the vibrational bands is complicated.

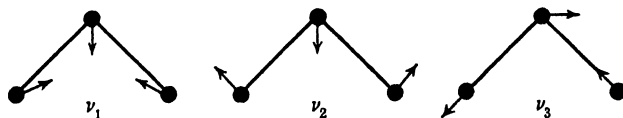


FIG. 125. Vibrational modes of nonlinear triatomic molecule

Electronic Spectra.—As already explained, the electronic spectra of molecules, appearing generally in the ultraviolet region, consist of a number of bands representing different vibrational transitions accompanying the electronic change; each band is made up of a number of fine lines due to changes in rotational energy superposed on the electronic and vibrational energy changes. On examination of a band group it is seen that each vibrational band is sharply defined at one end, called the **head** of the band, and diffuse at the other; the band is said to be "degraded" in the diffuse direction. The head of the band, although very useful for identification purposes, has no special significance; it is clearly marked because the rotational lines crowd together in its vicinity, but are much more widely separated at the degraded end. The "center" or **origin** of the band is sometimes marked by a gap, but more frequently it is occupied by a Q -branch, even with diatomic molecules.

If the difference in electronic energy between the initial and final levels may be expressed in the form $hc\bar{\nu}_{el.}$, then the frequency $\bar{\nu}$ cm.⁻¹ of any line resulting from the given electronic transition in a diatomic molecule is

$$\bar{\nu} = \bar{\nu}_{el.} + (v' + \frac{1}{2})\omega_e' - (v'' + \frac{1}{2})\omega_e'' + 2Bm, \quad (95)$$

assuming that the three different forms of internal energy are added algebraically, and that there is no interaction. In this equation v' and v'' are the vibrational quantum numbers in the initial and final electronic states, and ω_e' and ω_e'' cm.⁻¹ are the corresponding equilibrium vibration

frequencies of the molecule; B and m have the same significance as before, for they give the rotational fine structure of the vibrational band. A series of bands for which $v' - v''$ is constant is called a **sequence**, and the term **progression** is used to describe a set of bands having a definite value of either v' or v'' , while the other varies. It has been assumed in (95) that the oscillations of the molecule are harmonic, and that the moment of inertia is the same in the two electronic states; the latter assumption may introduce an appreciable error, but it is generally possible to make an empirical correction. Further, for an accurate analysis of the band spectrum it is necessary to allow for the anharmonicity; this may be done by introducing the additional term $-(v + \frac{1}{2})^2 x \omega_e$, as on p. 567.

From a careful study of the electronic spectrum of a gaseous substance, provided the complexities can be successfully unraveled, it should be possible to determine ω_e , x and B for the molecule. If the latter is diatomic, then these data may be used to calculate the moment of inertia, internuclear distance and force constant, as described on p. 569. The equilibrium frequencies can be obtained for both the ground state and the excited electronic states of the molecule, as given by (95), and if the correction for the change in moment of inertia is applied, the value of the moment may also be determined for both states. The electronic spectra have advantages over those in the infrared because they can be studied photographically with great accuracy; further, homopolar diatomic molecules, which do not interact with infrared radiation, give vibrational bands with rotational structure in the electronic spectrum. The alternation in intensity of the rotational lines has led to the discovery of ortho- and para-states (p. 94). It should be noted that all the vibrational modes discussed in connection with vibration-rotation spectra also appear in the electronic spectra. For polyatomic molecules these spectra are very complex and difficult to interpret.

Since the equilibrium vibration frequency ω_e depends on the nature of the isotopic atoms present in the molecule, it is evident from (95) that the vibrational bands produced by two isotopic forms of a substance will not coincide, for the origins will be slightly displaced. The calculation of the isotopic shift is somewhat more complicated than is the case with the vibrational spectrum, because there are four vibration frequencies to be considered, viz., ω' and ω'' for the two isotopic molecules. It is not a difficult matter to make the computation, and the observed results are in excellent harmony with those derived in this manner. As noted on p. 148, the isotopes of oxygen, carbon and nitrogen were discovered as the result of a study of electronic spectra.

It is convenient for many purposes to have a method of defining the electronic state of a molecule, and one similar to that described for atoms (p. 69) is employed. The resultant azimuthal quantum number for the molecule, which is exactly analogous to L for an atom, is represented by Λ , and the Greek letters Σ , Π and Δ are used as term symbols for $\Lambda = 0, 1$ and 2 , respectively. The symbol is preceded by $2S + 1$ as a superscript, where S is the resultant

spin number for the molecule. If the molecule consists of completed electron groups only, as is the case with most diatomic gases in their ground states, then $\Lambda = 0$ and $S = 0$, and the substance is in a $^1\Sigma$ state. The oxygen molecule is normally $^3\Sigma$, since it has two uncoupled electron spins (p. 107); S is then unity, although Λ is still zero.

Potential Energy Curves.—As a result of the attractive and repulsive forces between the atoms (cf. pp. 300, 407), the potential energy of a diatomic molecule may be represented as a function of the internuclear distance by a curve such as that in Fig. 126. Since the atoms are vibrating, the potential energy will oscillate about the minimum position M , and the vibrational levels may be indicated by the horizontal lines marked with the appropriate quantum numbers. Consider the line LN , for example; when the molecule vibrates the potential energy changes as the nuclei recede and approach, from L through M to N , and back again for every vibration. If the motion were simple harmonic the curve LMN would be symmetrical about M ; this is approximately true for small values of the quantum number, i.e., for small vibrations, but as the extent of vibration increases the anharmonicity factor becomes increasingly evident. It may be seen from Fig. 126 that as a result the nuclei should have infinite separation during the course of an oscillation when the vibrational quantum number reaches the value v_c ; that is to say, it should be possible to cause a molecule to dissociate by supplying it with sufficient vibrational energy. This type of dissociation is rare in the normal state of the molecule, but it may easily occur as the result of an electronic transition, as will be seen below.

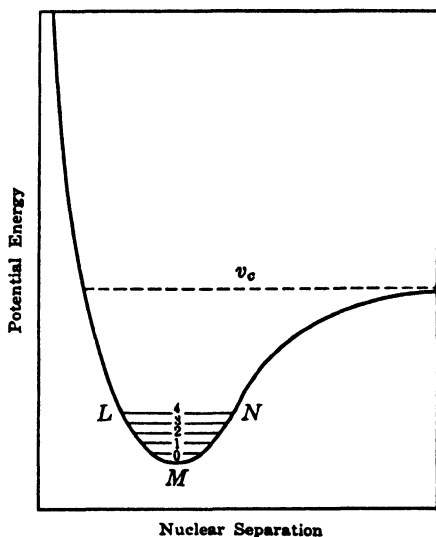


FIG. 126. Potential energy curve for diatomic molecule

The potential energy curve for a diatomic molecule may be represented, approximately, by the equation proposed by P. M. Morse (1929), viz.,

$$U(r) = D\{1 - e^{-a(r-r_e)}\}^2, \quad (96)$$

where $U(r)$ is the potential energy for an internuclear distance r , reckoned from the lowest point of the curve, M in Fig. 126, as zero; the equilibrium distance, i.e., at M , is r_e , and D is the dissociation energy, corresponding to the level v_c , also reckoned from the lowest point; a is a constant for the given molecule, equal to $(8\pi^2\mu x\omega_e c/h)^{1/2}$, where μ , x , ω_e , c and h are as previously defined. If r

in the Morse equation (96) is set equal to r_e , the value of $U(r)$ becomes zero, whereas as r approaches infinity $U(r)$ becomes equal to D , the dissociation energy, as required by Fig. 126. However as r approaches zero, the value of $U(r)$ remains finite, although it should increase to infinity. This weakness of the Morse function is not serious, because the conditions under which the internuclear distance approaches zero are of no practical significance.

Every electronic state has its own potential energy curve, the forms of which may be completely unrelated. The curves for a lower (*A*) and an upper (*B*) electronic level are shown in Fig. 127; the nuclear separations at equilibrium, i.e., for the minimum potential energy, are seen to be different, and so also are the energies of dissociation. The vertical distance between the minima of the two curves is equal to the difference in electronic energy, i.e., $\epsilon_{el.} = hc\bar{\nu}_{el.}$ (cf. p. 571), between the two states. According to J. Franck (1925) and E. U. Condon (1926), the transitions from one state to another are most probable when the nuclei are in their extreme positions, e.g., at *L* and *N* in Fig. 126, for it is there, when the vibrational kinetic energy is zero, that they spend the longest time.

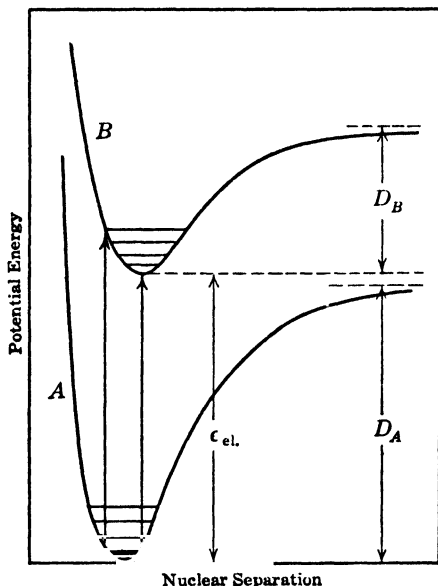


FIG. 127. Electronic transitions in diatomic molecule

Further, since the time required for an electronic transition is small compared with the time necessary for the nuclei to oscillate, the internuclear distance will remain constant and electronic transitions can, therefore, be represented by vertical lines, such as those shown in Fig. 127, starting from the extreme positions of an oscillation. According to quantum mechanics, the transition from the lowest ($v = 0$) level, however, takes place from the midpoint of the (classical) vibration, and not from the extremes. The foregoing conclusions are referred to as the **Franck-Condon principle**; they account for the changes observed in the vibrational quantum number in an electronic transition, and explain why there is no simple selection rule such

as is applicable to a purely vibrational transition.

In the electronic transitions depicted in Fig. 127 the molecule has a definite amount of vibrational energy in both upper and lower states; the spectrum corresponding to such transitions exhibits definite vibrational bands. If the potential energy curves of the two electronic states

are related somewhat in the manner of Fig. 128, certain transitions, such as the one shown, will be accompanied by dissociation. The vibrational energy in the upper state is seen to be so high that the internuclear distance will become infinite in the course of the vibration. When this type of electronic transition is possible, the absorption spectrum may consist of a number of bands, corresponding to transitions from the ground state to definite vibrational levels in the upper state, but these bands will converge and eventually pass into a continuous region, or **continuum**, showing no fine structure. In this region there is no complete vibration of the molecule in the upper electronic state, because the energy is sufficient to bring about dissociation. The energy of the **convergence limit**, that is, where the vibrational bands converge and become continuous, is a measure of the energy of dissociation of the molecule in the given state; as will be seen later, the products of dissociation are not necessarily normal atoms, but are frequently in electronically excited states. It may be mentioned that some molecules, e.g., hydrogen iodide, have upper electronic levels giving potential energy curves of the type of B' in Fig. 128; every transition from lower to upper state results in dissociation and no fine structure is observed.¹⁵

Heats of Dissociation.—It is only in a limited number of cases that the actual position of the convergence limit can be observed, so that the heat of dissociation may be calculated from it; some examples are given on p. 1165 in connection with the application of the results to photochemistry. If direct observation of the convergence limit is not possible, then approximate methods may be used to determine its position. It may be seen from (85) that the energy difference $\Delta\epsilon$ between two successive vibrational bands is

$$\Delta\epsilon = hc\omega_e - 2(v + \frac{1}{2})hc x\omega_e, \quad (97)$$

where ω_e is expressed in wave numbers. It is evident that as v increases $\Delta\epsilon$ becomes less and less, and finally at the convergence point, when v is equal to v_e , $\Delta\epsilon$ is zero. The value of v_e may be obtained graphically by plotting $\Delta\epsilon$ against v , from the observed separations of the origins of successive vibrational bands, and extrapolating to $\Delta\epsilon$ equals zero. The area under the curve between $v = 0$ and $v = v_e$ gives the difference in energy between the lowest level and the convergence limit, and this quantity, represented by D_0 is equal to the heat of

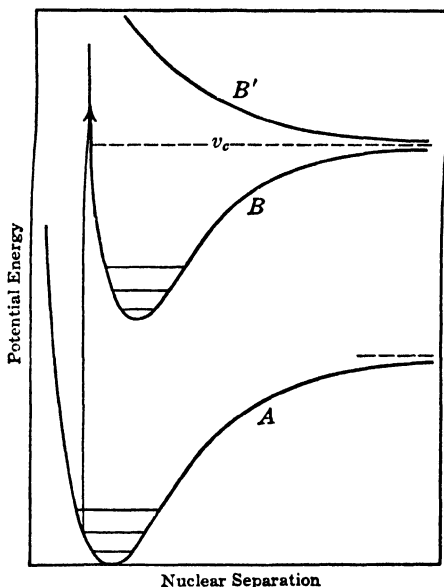


FIG. 128. Electronic transition leading to dissociation

dissociation of the molecule. Alternatively, if v in (97) is replaced by v_e , and the expression equated to zero, it follows that

$$v_e = \frac{1}{2}(1 - x)/x, \quad (98)$$

and if this is substituted in (85), the vibrational energy at the convergence limit, which corresponds to the energy previously represented by D (p. 573), for a single molecule, is given by

$$D = \epsilon_e = \frac{1}{4}hc\omega_e/x. \quad (99)$$

The heat of dissociation is obtained by subtracting from this the energy of the molecule in its lowest energy level, when $v = 0$ and $J = 0$, that is, the **zero-point energy** (p. 157); this may be readily evaluated by putting $v = 0$ in (85), when it is seen that

$$\epsilon_0 = \frac{1}{2}hc\omega_e - \frac{1}{4}hc x \omega_e,$$

per single molecule. For most purposes $x\omega_e$ may be neglected in comparison with ω_e , so that

$$\epsilon_0 = \frac{1}{2}hc\omega_e. \quad (100)$$

The heat of dissociation is then

$$D_0 = D - \epsilon_0 = \frac{1}{4}hc\omega_e \left(\frac{1}{x} - 2 \right) \quad (101)$$

per single molecule; this must be multiplied by the Avogadro number to obtain the heat of dissociation per mole. Since ω_e and x for the ground state of the molecule can be obtained from an analysis of the band spectra, the value of D_0 may be calculated. The results obtained are not always accurate, the chief source of error being the fact that the energy of a vibrating molecule is not given exactly by (85), and so there may thus be a serious discrepancy in the estimate of v_e as expressed by (98). The method of graphical extrapolation is generally more reliable, but the difficulty here is that the extrapolation has to be carried over such a long range that the slope of the curve becomes uncertain.¹⁶

Raman Spectra.—A valuable adjunct to electronic and infrared spectra is the Raman effect, predicted theoretically by A. Smekal (1923) and discovered by C. V. Raman (1928). If any substance, gaseous, liquid or even solid, is exposed to radiation of a definite frequency, then the light scattered at right angles contains frequencies, differing from that of the incident radiation, which are characteristic of the substance under examination. If $\bar{\nu}_i$ is the frequency of the incident light and $\bar{\nu}_s$ that of the scattered light, then the difference, $\bar{\nu}_i - \bar{\nu}_s = \Delta\bar{\nu}$, is called the Raman frequency. The accepted mechanism of the Raman effect is that a molecule absorbs the incident radiation and is thereby raised to a higher level; it then emits the Raman scattered radiation and falls to a level usually between the initial and excited states. The energy $hc\Delta\bar{\nu}$ is then the difference between initial and final states. In some instances the final level has less energy than the initial value, so that $\bar{\nu}_s$ is greater than $\bar{\nu}_i$; the radiations so produced are called "anti-Stokes" lines (cf. p. 1155).

An examination of the Raman spectra of many compounds has shown that $hc\Delta\bar{\nu}$ is almost invariably equal to the change in rotational or vibrational energy of the molecule; the latter type of spectra are the more important, and the information obtained from them has helped in the

elucidation of electronic and vibrational bands of the normal type. The study of the Raman effect has many advantages. By appropriate choice of the incident radiation the scattered lines are brought into a convenient region of the spectrum where they can be easily observed. Homopolar diatomic molecules give Raman spectra due to vibrational and rotational transitions although they do not interact directly with infrared radiation, and vibrations which are inactive in normal spectra are often active in the Raman effect, e.g., the ν_1 frequency of carbon dioxide (p. 570). In general, a vibration is active in the Raman spectrum when it is accompanied by a change in the polarizability of the molecule. Further, a Raman line indicating a Q -branch, corresponding to the origin of a vibrational band, often appears where none is found in the vibration-rotation spectrum. It is evident, therefore, that the two methods of studying vibrational energy changes are complementary; when vibrations are active in both types of spectra, the agreement between the results is very satisfactory. Because of the ease with which Raman spectra may be studied a large amount of work has been done in this field.¹⁷

Applications of Molecular Spectra: Force Constants.—For a diatomic molecule, treated as a harmonic oscillator, the restoring force per cm., i.e., the force constant, can be determined from the known value of the equilibrium frequency by means of (94). In this case, combination of the expression for the potential energy, i.e., $\frac{1}{2}fa^2$, where a is the displacement, with one based on mechanical considerations, leads to an equation, i.e., (94), containing one unknown, the force constant, which can be calculated from the vibration frequency derived from a study of the spectrum. With polyatomic molecules, however, the situation is rendered difficult by the fact that there are more unknowns, the force constants, than there are knowns, the frequencies. In general, the potential energy of vibration may be written

$$U = \frac{1}{2}(f_1a_1^2 + f_2a_2^2 + \cdots + 2f_{12}a_1a_2 + 2f_{23}a_2a_3 + \cdots), \quad (102)$$

where f_1, f_2 , etc., are the restoring forces, and f_{12}, f_{23} , etc., are interaction terms between the various forces; a_1, a_2 , etc., are the displacements of the various atoms in the vibration. The potential energy can also be expressed in terms of the vibration frequencies, but as there are only $3n - 6$ vibrations in the molecule it is impossible to solve (102), since this contains $\frac{1}{2}n(n - 1)$ force constants in addition to the interaction terms. In order to make progress it is necessary to reduce (102), by making approximations, until it contains not more than $3n - 6$ terms; a solution of the equation is then, theoretically, possible. The simplest assumption, which has proved very successful, is that known as the **valence force field**: a force constant is associated with every valence bond, representing its resistance to stretching, and to every bond angle, which measures its resistance to deformation. Equation (102) may then be written

$$U = \frac{1}{2}(f_1\Delta d_1^2 + f_2\Delta d_2^2 + \cdots + f_{\theta_1}\Delta\theta_1^2 + f_{\theta_2}\Delta\theta_2^2 + \cdots), \quad (103)$$

the interaction terms being omitted; in (103), Δd_1 , Δd_2 , etc., are the displacements of the nuclei in the direction of the valence bonds, and $\Delta\theta_1$, $\Delta\theta_2$, etc., are the changes in the angles. In general, the number of unknowns is reduced to such an extent that a solution is possible with the aid of an equivalent number of equations derived from purely mechanical considerations of the molecular oscillations. In this way the force constants for various bonds in a number of molecules have been determined; apart from the difficult problem of assigning the correct frequencies to the different vibrations, the actual calculations are by no means simple.

The force constant of a given bond is almost constant although, as might be anticipated, the remainder of the molecule has some influence; the existence of resonance, which alters the nature of the bond (p. 111), also has an effect, and in fact the force constant may be utilized, as indicated below, to detect this phenomenon. Some mean values for the force constants of single, double and triple bonds are quoted in Table 92.

TABLE 92. FORCE CONSTANTS OF VARIOUS LINKAGES IN DYNES PER CM.

Bond	Force Constant	Bond	Force Constant	Bond	Force Constant
C—C	4.6×10^5	C—O	4.9×10^5	C—N	4.8×10^5
C=C	9.5	C=O	12.3	C=N	12.1
C≡C	15.8	C≡O	18.6	C≡N	17.5

It is an interesting fact that the force constants increase approximately in proportion to the multiplicity of the bond, and so the former can be used to give an indication of the latter. The force constant given for C≡O is that for carbon monoxide, and the value, 18.6×10^5 dynes per cm., establishes without doubt the important contribution of the triple bonded structure. In carbon dioxide the force constant for the carbon-oxygen bond is 15.2×10^5 dynes per cm. This result is in agreement with the resonance structure described on p. 112; the bonds evidently have both double and triple bond character.

Characteristic Frequencies.—The vibration frequencies determined from spectra are those of the different modes of vibration of the molecule as a whole. It has become evident, however, that the force constants of certain linkages are often the factors which exercise the main influence on a particular frequency. In fact this frequency may be regarded as being confined to the controlling bond, and it has become the practice to associate a characteristic frequency with each type of linkage. It must be emphasized that this is only an approximation, but it is one which has proved useful. For the identification of these frequencies the Raman spectra have been most generally employed, for the difference in frequency of the incident and scattered radiation gives directly the approximate vibration frequencies of the oscillators in the molecule. The results in Table 93 have been mostly obtained from Raman spectra; the actual frequencies depend to some extent on the remainder of the molecule, and so are only approximate. According to (94) the vibration frequency

TABLE 93. CHARACTERISTIC FREQUENCIES OF LINKAGES

Bond	Frequency	Bond	Frequency	Bond	Frequency
C—C	800– 860 cm^{-1}	C—O	820– 880 cm^{-1}	C—N	880–930 cm^{-1}
C=C	1600–1650	C=O	1710–1750	C=N	1650
C≡C	2100–2250	C=O	2160	C≡N	2150

of a diatomic molecule is related to the restoring force, and a similar general connection is evident from comparison of Tables 92 and 93. It will be seen that single bonds involving carbon, nitrogen and oxygen have characteristic frequencies about 800 cm^{-1} , double bonds about 1600 cm^{-1} , and triple bonds approximately 2100 cm^{-1} . If hydrogen is attached to one of these atoms, e.g., C—H, N—H or O—H, the frequency increases to about 3000 cm^{-1} , and the force constant increases correspondingly. With heavier atoms the frequency diminishes, e.g., 650 to 710 for C—Cl, as may be expected from general considerations, since the reduced mass is greater. Data of the type shown in Table 93 can also be used to provide information concerning molecular structure. For example, the characteristic frequency of the isocyanide group is about 2150 cm^{-1} , showing that the triple bonded structure $-\text{N}\equiv\text{C}-$ must predominate. The carbon-carbon frequency in an aromatic ring is about 1600 cm^{-1} , in harmony with the resonance formulation (p. 112).

Interatomic Distances.—Since moments of inertia can be obtained from the separation of the rotational lines, the study of spectra has provided accurate information on molecular dimensions. Some of the results, for hydrogen halides, have already been given on p. 569, where the method of calculation has been indicated. For more complex molecules, involving three moments of inertia and several interatomic distances, the computation is naturally more difficult. Interatomic distances involving hydrogen atoms cannot be determined accurately by other methods, and the values derived from the spectra of methane, ammonia and water are of particular interest; they are as follows:

C—H	N—H	O—H
1.09 Å.	1.01 Å.	0.96 Å.

From the moments of inertia of ethylene and acetylene the C=C and C≡C distances have been estimated as 1.33 Å. and 1.20 Å., respectively; these have been confirmed by electron diffraction measurements (p. 591). Since the C—C distance is 1.54 Å., as found in diamond and many organic compounds, the double-bond length is about 87 per cent and the triple-bond length about 78 per cent of the single bond distance. These proportions are assumed to apply to linkages between carbon and other atoms, e.g., oxygen and nitrogen. If, as a result of resonance between two or more different structures, a bond has mixed character, e.g., mixed single and double carbon-carbon bonds in benzene, the bond length is intermediate between those for pure single and double bonds, viz., 1.39 Å.

in benzene. In general, the length tends to be closer to that of the higher multiple bond (cf. p. 112).¹⁸

As seen above, the distance between two linked atoms depends on the nature of the linkage, and a connection also exists between the latter and the force constant (Table 92). A quantitative relationship between bond lengths and force constants is thus to be expected. R. M. Badger (1934) proposed the equation

$$r = C/f^{\frac{1}{2}} + d_{ij}, \quad (104)$$

where r is the equilibrium bond distance and f is the corresponding force constant; C is almost constant for all bond types, while d_{ij} is a constant whose value depends on the series in the periodic table containing the two bonded atoms. For carbon-carbon bonds, for example, the Badger equation may be written $r = 0.571/f^{\frac{1}{2}} + 0.680$, and the same constants apply approximately to carbon-oxygen and carbon-nitrogen bonds. Another relationship, which has proved satisfactory in some cases, is that of C. H. D. Clark (1935), viz.,

$$r = (C\mu/nf)^{1/6}, \quad (105)$$

where C is a constant for elements in a given period, and n is the sum of the valence electrons of the bonded atoms; μ is the reduced mass of the two atoms. This equation is, however, fundamentally unsound, because a change in μ , resulting from an isotopic exchange which leaves the force constant unaffected, should result in an alteration in the bond length; this is apparently in disagreement with the experimental facts. A modification, suggested by H. S. Allen and A. K. Longair (1935), which may be written in the form

$$r = C/f^{1/6}, \quad (106)$$

overcomes this objection. It should be noted that the foregoing empirical equations are approximations to a theoretical relationship which undoubtedly exists between force constants and bond lengths.¹⁹

SPECTRA OF LIQUIDS

Spectra of Liquids: Infrared Spectra.—As a consequence of molecular interaction, rotational fine structure is not observed in the spectra of liquids. Vibrational bands and Raman spectra have been studied but the data are not of such quantitative value as those obtained from gases and vapors. It is nevertheless possible to associate certain absorption bands in the infrared with the vibrations of specific groups in the molecule: for example, aliphatic compounds generally give a band with an absorption maximum at about 3.4μ and a harmonic at 1.7μ , which are undoubtedly connected with the C—H linkage. The position of the band may vary from one substance to another, because of the influence of the remainder of the molecule. In aromatic compounds the C—H band appears at 3.25μ . Many hydroxylic compounds exhibit a sharp absorption maximum at about 2.75μ and a broader band near 3μ ; the former is strong in the vapor of ethyl alcohol, but weak in the liquid at ordinary temperatures. The second band is strong in the latter, but is weakened

if the alcohol is diluted with an inert liquid, or if its temperature is raised. Similar observations have been made with other hydroxylic compounds, and the conclusion drawn is that the band at 2.75μ is due to the simple O—H group, whereas the other is produced when the hydrogen atom takes part in the formation of a hydrogen bond (p. 509). Measurements of infrared spectra are providing a valuable method for the study of association and complex formation, involving hydrogen bridges, in pure liquids and in solution.

Electronic Spectra of Liquids.—The presence of vibrational bands in the electronic spectra has been detected in relatively few cases, e.g., benzene, in which the electrons in the molecule may be regarded as having a closed path. For most liquids the ultraviolet absorption spectrum appears in the form of one or more bands of apparently continuous absorption. The failure to observe the vibrational structure may sometimes be due to the inadequate resolving power of the spectroscope, but it is more often caused by the overlapping of several vibrational bands; further, interaction between neighboring molecules results in a variation of the vibrational energy so that the effect of quantization is destroyed. In the ultraviolet spectrum of liquid benzene there are eight vibrational bands in the region between 36,000 and 44,000 cm^{-1} , i.e., 2270 to 2780 \AA . wave length, but this structure disappears to a great extent, or completely, in derivatives of benzene. In spite of the absence of definite structural features, electronic absorption spectra in the ultraviolet region have proved of interest and practical value, because compounds containing the same characteristic groups give analogous spectra, either in the pure liquid state or when dissolved in a given solvent. Application of this general rule, with certain reservations which will appear later, has led to the elucidation of the structures of a number of compounds.

Extinction Coefficients: Beer's Law.—For the recording and comparison of absorption spectra of pure liquids and solutions, it is the practice to plot the "extinction coefficient" against the wave length, or wave number; in this way a curve characteristic of the light-absorbing substance is obtained. Light is passed through the liquid contained in a vessel with quartz faces, and the emergent beam enters a quartz-prism or a grating spectrograph. By a suitable photometric analysis of the resulting spectrum the extent of light transmission at each wave length or frequency can be determined. If I is the intensity of the light transmitted at a given wave length and I_0 is that of the incident light, then according to Beer's law,

$$I = I_0 10^{-\epsilon cd},$$

where c is the concentration of the solution in moles per liter and d is the thickness through which the light is transmitted; ϵ is known as the **molar extinction coefficient** of the solute for the particular wave length. Sometimes Beer's law is used in the form $I = I_0 e^{-\alpha cd}$, where α is called the

molar absorption coefficient, and for a pure liquid Lambert's law, viz., $I = I_0 10^{-kd}$, k being the **extinction coefficient** of the substance. The plot of the appropriate extinction coefficient against the wave length gives an "absorption curve" characteristic of each substance. If the values of the extinction coefficient are large, that is to say, when there is strong absorption in certain regions of the spectrum, it is more convenient to plot $\log \epsilon$, as shown in Fig. 129. The most intense absorption is seen to be

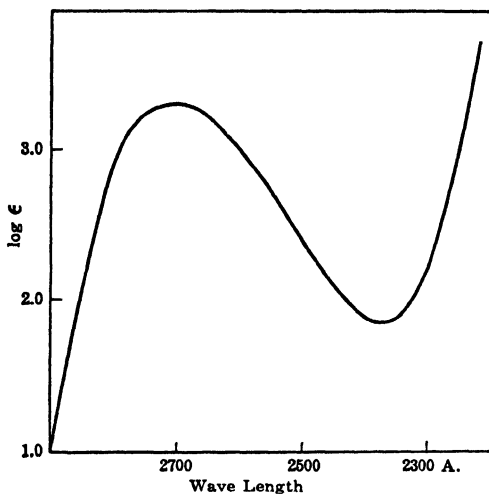
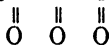


FIG. 129. Absorption curve

in the vicinity of 2700 Å. wave length, and hence the substance is said to have an **absorption band** at 2700 Å. Some compounds exhibit two or more bands of varying intensities. The possible effect of the solvent on the absorption curve will be considered shortly.²⁰

Color in Organic Compounds.—The absorption spectrum of a substance is, of course, related to its color; when the absorption is in the ultraviolet only, the substance will be colorless, but if structural changes in the molecule cause the band to shift to regions of longer wave length, visible color can result. The longer the wave length of the absorbed radiation, i.e., the smaller the frequency, the smaller is the energy change involved in the associated electronic transition. In completely saturated aliphatic compounds the electrons are tightly bound, and hence a large amount of energy would presumably be required to raise the molecule to an excited state. Such substances exhibit absorption in the far ultraviolet region of the spectrum. If unsaturated linkages, in the form of such groupings as $-\text{C}=\text{C}-$, $-\text{N}=\text{N}-$, $-\text{C}=\text{O}$, $-\text{N}=\text{O}$, etc., frequently called **chromophores** (O. N. Witt, 1876), are present in the molecule, the p_z (or π) electrons (see p. 105), sometimes referred to as **unsaturation electrons**, are less firmly bound than those forming normal single linkages. Less energy is therefore required for an electronic transition, and hence the absorption band appears in the near ultraviolet region, and sometimes in the visible

spectrum. Two or more unsaturated groupings often enhance one another so that visible color is produced, as in dimethyl keten, i.e., $(\text{CH}_3)_2\text{C}=\text{C}=\text{O}$, and in the triketones, e.g., $\text{CH}_3\text{C}-\text{C}-\text{CCH}_3$; this is especially true if the



double bonds are conjugated, as in the latter compound and in phorone, $(\text{CH}_3)_2\text{C}=\text{CH}-\text{C}-\text{CH}=\text{C}(\text{CH}_3)_2$. The general phenomenon of the shift of



an absorption band to a region of longer wave length is often referred to as **bathochromic**; hence, conjugation of double bonds has a bathochromic effect.

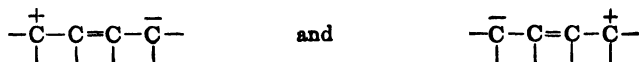
As might be expected from its electronic structure, benzene, with its six p , electrons (p. 112), shows appreciable absorption in the near ultraviolet; the introduction of a conjugated unsaturated side chain produces a shift to longer wave lengths. The diphenyl polyenes, $\text{C}_6\text{H}_5(\text{CH}=\text{CH})_n\text{C}_6\text{H}_5$, form an interesting series in this connection; stilbene itself, $\text{C}_6\text{H}_5\text{CH}=\text{CHC}_6\text{H}_5$ is colorless, but as the number of $-\text{CH}=\text{CH}-$ groups is increased, the color passes through yellow to green and orange, and when n is 8 the compound has a bluish coppery-red color. The introduction of a $-\text{CH}_2$ group in the unsaturated chain destroys the conjugation, and is accompanied by a shift of the absorption band to shorter wave lengths; that is to say, the $-\text{CH}_2$ group has a **hypsochromic** effect. This phenomenon may be illustrated by reference to $\text{C}_6\text{H}_5(\text{CH}=\text{CH})_4\text{C}_6\text{H}_5$, which is greenish yellow, while $\text{C}_6\text{H}_5\text{CH}_2(\text{CH}=\text{CH})_4\text{C}_6\text{H}_5$ is colorless. An important application of the influence of increasing conjugation is found in the cyanine dyes; these consist of two heterocyclic nuclei joined by a $=\text{CH}(\text{CH}=\text{CH})_n-$ chain. As the value of n increases the absorption band of the dye moves from the violet region to the red and near infrared. It is the property of being able to absorb in the latter part of the spectrum which makes certain cyanine dyes particularly useful for rendering photographic plates sensitive to infrared rays (see p. 1187). The $-\text{N}=\text{N}-$ group behaves similarly to the $-\text{CH}=\text{CH}-$ grouping in some ways. For instance, $\text{CH}_3\text{N}=\text{NCH}_3$ has a relatively feeble absorption band at 3450 \AA ., but $\text{C}_6\text{H}_5\text{N}=\text{NC}_6\text{H}_5$, in which conjugation may be supposed to continue through both phenyl groups, absorbs strongly at 4450 \AA . and so is visibly colored; if, however, one or two $-\text{CH}_2$ groups are introduced, e.g., as in $\text{C}_6\text{H}_5\text{CH}_2\text{N}=\text{NCH}_2\text{C}_6\text{H}_5$, the absorption band, now weakened considerably, appears in the short wave length region.

It has been mentioned above that one chromophoric grouping often enhances another in the same molecule; this is only true if the groups are relatively close together or are separated by a conjugated system across which electronic shifts may be transferred. If the two chromophores are separated by a saturated chain of three or more carbon atoms, e.g., as in $\text{C}_6\text{H}_5\text{CH}_2\text{CH}_2\text{CH}_2\text{I}$, each exerts its own effect, so that the absorption spectrum of ω -phenyl propyl iodide is almost the same as those of propyl benzene and propyl iodide superimposed on one another. Similarly, $\text{C}_6\text{H}_5\text{CH}_2\text{CH}_2\text{CHO}$ gives the same absorption spectrum as a mixture of benzene and propionaldehyde.

Free radicals and odd-electron molecules are usually associated with absorption bands in the visible spectrum; this is to be expected from the fact that one of the electrons is probably loosely bound, although other factors are of importance.

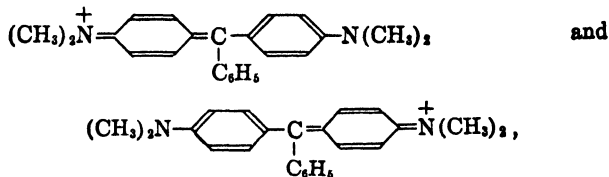
In dyestuffs the intense color is associated with the presence of a chromophoric structure, such as a conjugated chain or one or more benzene rings, and of an **auxochromic group** or groups. It is the combination of the two groupings which produces strong absorption of light in the visible spectrum. A simple example is provided by phenol in alkaline solution; the ion $\text{C}_6\text{H}_5\text{O}^-$ is colorless, but the introduction in the *para*-position of the $-\text{NO}_2$ group, acting as an auxochrome, produces the yellow $p\text{-NO}_2\text{C}_6\text{H}_4\text{O}^-$ ion. A further illustration is the triphenylmethyl cation, which is colorless or pale yellow in acid solution, whereas the corresponding ion containing two or three hydroxyl groups in the *para*-positions is deep red. Auxochromic groups can apparently be divided into two categories: the basic (or positive) groups, e.g., $-\text{OH}$, $-\text{OR}$, $-\text{NR}_2$, $-\text{NHPh}$, $-\text{NH}_2$, are usually effective in acid solution, while the acidic (or negative) groups, e.g., $-\text{NO}_2$, $-\text{NO}$, $-\text{CO}$, $-\text{CN}$, function in alkaline solution. The aforementioned auxochromes refer particularly to dyes in which the benzene ring is part of the chromophore; substitution in the *ortho*- and *para*-positions, especially the latter, gives the greatest intensification of color, but *meta*-substitution has little or no influence.

Resonance and Color.—There is no doubt that the presence of color in organic compounds is to be attributed to the possibility of resonance; further, intense color is usually observed when the two most important resonance forms are such as to involve the movement of an electric charge over a relatively large distance within the molecule. In a system of conjugated double bonds, for example, the two extreme ionic structures which are involved in resonance, in addition to the normal structure, $-\text{C}=\text{C}-\text{C}=\text{C}-$, are



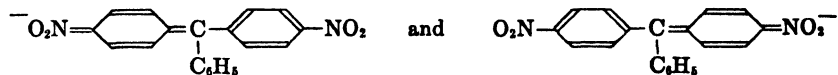
There are also intermediate ionic structures, in which the separation of the charges is not so great; the number of such possible structures involved in resonance increases rapidly as the number of conjugated double bonds increases. This may be correlated with the observations on the colors of the polyenes mentioned earlier.

The function of an auxochrome group in a dye is to contribute to the resonance structures already present and to provide new ones. Consider, for instance, the pale-colored triphenylmethyl cation, $(\text{C}_6\text{H}_5)_3\text{C}^+$, in acid solution; the resonating forms are those usually associated with the benzene rings, giving absorption in the near ultraviolet, but no visible color. The introduction of two $-\text{N}(\text{CH}_3)_2$ groups in the *para*-positions gives the dye malachite green, while three such groups produce crystal violet. In the former case the two extreme (identical) resonating structures are



and analogous structures can be written for the latter dye. In each substance the resonance forms differ by the movement of an electric charge over a con-

siderable distance. The introduction of the auxochrome groups thus makes possible new resonance forms, in addition to those already present in the $(C_6H_5)_3C^+$ ion. It will be noted that similar resonance structures can be written when the $-N(CH_3)_2$ groups are in the *ortho*-positions, but not when they are in the *meta*-positions. Further, the possibility of stable resonance structures depends on the presence of the positive charge, that is, on the fact that the solution is acid; this explains why the $-N(CH_3)_2$ group is a basic (positive) auxochrome. In alkaline solution substitution of this group in the triphenylmethyl anion $(C_6H_5)_3C^-$ would not yield a dye, since resonance of the type just described is not possible; however, the introduction of the acidic (negative) $-NO_2$ groups produces intense color. The extreme resonance structures can now be represented as



The effectiveness of other acidic and basic auxochrome groups can be accounted for in a similar manner.

The exact reason why resonance produces intense absorption at long wave lengths is not entirely clear, although a number of theories have been proposed (R. S. Mulliken, 1939; L. Pauling, 1939). One suggestion is that the large number of intermediate ionic forms contribute mainly to resonance in the excited state, thus resulting in a marked lowering of the energy in the latter state (cf. p. 111); the energy required to raise the molecule from its ground state to the excited state is then relatively small, and so absorption occurs at long wave lengths. Another possibility is that the combination of the resonating forms of the molecule results in a splitting of the ground state into two levels which do not differ very greatly in energy.* The transition from the lower to the upper of these electronic levels would give an absorption of light of relatively long wave length. The extent of absorption, that is, the magnitude of the extinction coefficient, as distinct from the wave length of the band, should be dependent on the change of dipole moment of the molecule in passing from the ground state to the excited state. Movement of an electric charge over a considerable distance in the extreme resonating forms might thus be expected to give intense color, as has been actually observed.

A quasi-classical point of view, which has provided an interpretation of many absorption spectra phenomena, has been described by G. N. Lewis and M. Calvin (1939). Excitation of a colored molecule by radiation is supposed to produce a quantized oscillation of the conjugated system of mobile (unsaturation) electrons, which move in unison from one extreme resonating structure to the other; the motion may then be treated as equivalent to a single classical electronic oscillation within the molecule. According to classical electromagnetic theory, such an oscillation could result from the interaction of the electric vector of the absorbed light with a regularly reversing electric dipole within the molecule. The electronic oscillations, which will correspond in frequency to that of the absorbed radiation, will take place preferentially in the direction of maximum polarizability, but oscillations, associated with less intense absorption, are possible in the two other polarizability directions at

* The situation is somewhat analogous to that resulting in the formation of the two forms of the hydrogen molecule, with symmetric and antisymmetric orbitals, respectively (p. 102).

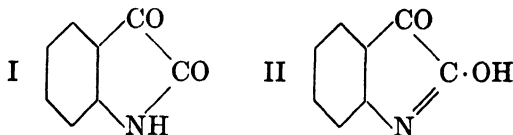
right angles. A flat symmetrical molecule, such as crystal violet, is equally polarizable in all directions in the plane, and so it exhibits one strong absorption band, at 5900 Å. This is known as the x band. In malachite green (see above), however, the polarizability is not uniform. The direction of maximum polarizability is obviously that from one $(\text{CH}_3)_2\text{N}-$ group to the other; oscillation in this direction is similar to that in crystal violet, and so malachite green has an intense x band at 6250 Å. The electronic oscillation in the less polarizable direction at right angles then leads to the weaker y band, observed at 4230 Å. In general both x and y bands have been detected for unsymmetrical planar molecules, but x bands only for symmetrical structures. Further, according to the theory, higher excitation energy should lead to a harmonic oscillation with a frequency approximately twice that of the fundamental. Second order bands, called x' bands, corresponding to these oscillations, have been found in a number of cases.

In a long conjugated chain, such as occurs in the diphenyl polyenes, the polarizability is virtually restricted to the direction of the chain; the intensity of absorption and the wave length of the band would thus be expected to increase with increasing chain length, as is actually observed. The plot of the number of $-\text{CH}=\text{CH}-$ groups against the square of the wave length of the absorption band was found to be a straight line, in agreement with the idea of electronic oscillation.²¹

Influence of Solvent.—The nature of the solvent often has an influence on the position of an absorption band. This fact is at once evident from a consideration of the colors of solutions of iodine in various liquids; in nonpolar solvents, e.g., *n*-hexane, carbon tetrachloride and carbon disulfide, the absorption maximum occurs at almost the same wave length as in iodine vapor, viz., 5180 Å. In polar solvents, however, a brownish color replaces the usual purple color of iodine; the absorption now occurs at shorter wave lengths. The change is sometimes attributed to polarization of the iodine molecule by the electric field of the solvent dipoles, but it is probable that in certain cases complex formation occurs. The shift in the position of an absorption band depends on the type of solvent and solute, and the many and varied observations are not easy to summarize. The following rules may be regarded as rough generalizations, although they are not universally applicable. If the absorbing substance is nonpolar, the effect of the solvent is determined chiefly by the refractive index of the solvent, and the same is (approximately) true for polar and ionic light absorbers in nonpolar solvents. For the two latter types of light absorbing species in polar solvents, however, the dipole moment of the solvent appears to be the factor which determines the change in the position of the absorption band. The formation of a compound between solvent and solute invalidates the foregoing rules, for the band may then appear in an entirely new position.

Applications of Absorption Spectra.—One of the best known of the early applications of absorption spectra was to the determination of the structure of isatin, for which two formulae, viz., lactam (I) and lactim

(II), had been proposed.



The two possible methyl ethers are known, and the spectrum of isatin itself is very similar to that of the N-methyl ether, thus favoring the lactam type of structure I (W. N. Hartley and J. J. Dobbie, 1899). The phenolic structure of phloroglucinol was proved in an analogous manner, for its ultraviolet spectrum resembles closely that of its methyl ether, which can only be derived from the phenolic form. The fact that styrene and tetrahydrobiphenyl have similar absorption spectra shows that the latter is phenyl-cyclohexene, with the double bond conjugated to those in the benzene ring, as in styrene (p. 530). Studies of ultraviolet absorption spectra have also been employed to throw light on the structures of alkaloids and of compounds belonging to the purine group.

The spectra of the metallic derivatives of acetyl acetone are of the same type as that of the parent diketone in alcoholic solution indicating, in agreement with modern views, that they have analogous structures derived from the enolic form. With acetoacetic ester, however, the results are different. The absorption spectra of the metallic derivatives are similar to those of acetyl acetone and its derivatives, but are not the same as that of the parent substance. This result is in harmony with the accepted view that ethyl acetoacetate exists mainly in the keto-form (see p. 531). In the presence of alkali the enolic form is produced and the spectrum changes accordingly. By observations on the ultraviolet spectrum, A Hantzsch (1910) was able to study the influence of temperature, dilution and nature of the solvent on the keto-enol equilibrium of acetoacetic ester. Other isomeric changes, such as that between nitrocamphor and the *pseudo*-compound, have been followed by means of the absorption spectra of solutions (T. M. Lowry and C. H. Desch, 1909).

The two forms of an acid-base indicator obviously have different absorption spectra; one has a strong band at a wave length at which the other exhibits no appreciable absorption. The extinction coefficient in this band, for a given total concentration of indicator, will thus depend on the relative amounts of the two forms of the indicator and hence on the acidity of the medium. The spectrophotometric method, in conjunction with suitable indicators, thus forms an accurate means for the study of hydrogen ion concentrations (W. R. Brode, 1924). Although producing no visible color, it is possible to study, by an analogous procedure, the extent of salt formation of weak organic bases by examining the ultraviolet absorption spectra in the presence of various proportions of a strong acid.²²

Measurements of absorption spectra find a number of other analytical applications, particularly in biological work where only small quantities of material are available. An amount sufficient to make a few cubic centimeters of a solution of 0.001 M, or less, is adequate for the purpose of plotting the absorption curve; by comparison of the extent of light absorption in a given solution with that of a standard, the concentration of the former can be evaluated.

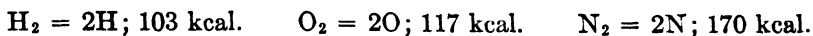
Absorption spectra have been utilized in connection with various problems in inorganic chemistry. For example, it has been established that the deep red color of ferric thiocyanate solutions is due to FeCNS^{++} ions, and the proportions of these ions present in various mixtures of ferric and thiocyanate ions have been determined from extinction coefficient measurements (H. E. Bent and C. L. French, 1941). It should be noted that it is the FeCNS^{++} ion and not the molecule of $\text{Fe}(\text{CNS})_3$, as was at one time supposed, that is responsible for the red color. The structure of the colored substance formed by ferric chloride with phenolic compounds has been investigated in an analogous manner (E. F. Wesp and W. R. Brode, 1934). Use has also been made of absorption spectra to explain the color changes from red to blue, and the reverse, observed in solutions of cobalt halides, and to study the structural changes occurring on dilution of nitric acid solutions.²³

THERMOCHEMICAL QUANTITIES

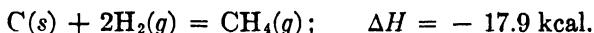
Determination of Bond Energies.—The heats of combustion, and hence of formation, of isomeric organic compounds are almost the same, and in any homologous series the increase per CH_2 group is almost constant at about 160 kcal.; hence it appears that heats of formation associated with particular bonds are to a great extent additive. This fact has been utilized to calculate the quantity known as the **bond energy**, associated with various linkages. The bond energy, sometimes called the heat of formation of the bond, is the average amount of energy required to dissociate bonds of the same type in 1 mole of a given compound. For example, one-fourth of the energy necessary to dissociate 1 mole of methane into atomic carbon and hydrogen is taken as the bond energy of the C—H linkage. This is not necessarily the same as the heat required to remove a single hydrogen atom from a molecule of methane; the latter quantity has been referred to as the **bond strength**.* More energy is required to dissociate the first hydrogen atom than for the removal of successive atoms from methane, and the bond energy represents a mean of the four bond strengths. In general, therefore, the "strength" of a particular bond will depend on its position or type in a given molecule, but the bond energy is assumed to be constant, with some justification, as will be seen later.

* Bond strengths have been estimated in a number of cases from studies of the kinetics of chemical reactions.

The procedure for the calculation of bond energies was indicated by K. Fajans (1920); it involves a knowledge of the heats of dissociation into atoms of the common elements, e.g., hydrogen, carbon, nitrogen and oxygen, and these will be considered first. The values for hydrogen, nitrogen and oxygen are known with fair accuracy from heats of dissociation of the corresponding molecules, determined spectroscopically; where a definite convergence limit can be observed this is an accurate method, and the results may be accepted with confidence. The heats required to split hydrogen, oxygen and nitrogen molecules into atoms are as follows:



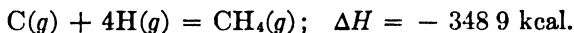
The heat of atomization of carbon is very uncertain and values ranging from 110 to 170 kcal., based on spectroscopic measurements, heat of sublimation, etc., have been proposed; most authorities prefer the value of 125 kcal. for the process $\text{C}(\text{solid}) \rightarrow \text{C}(\text{vapor})$, and this figure will be employed here. From the heats of combustion of methane, carbon and hydrogen, the heat of formation of methane is known; thus,



and introducing

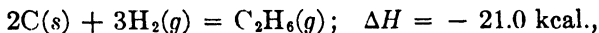


it follows that

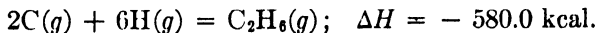


Since the formation of methane involves four C—H linkages, it appears that the mean heat of formation of each linkage, that is, the C—H bond energy, is 87.2 kcal.

The heat of formation of ethane is given by



and utilizing the heats of formation of carbon and hydrogen atoms quoted above, it follows that



Since ethane contains six C—H bonds, the heat of formation of which is 523, the energy of the remaining C—C linkage is 57 kcal. In the same manner it is possible to evaluate the bond energies of $\text{C}=\text{C}$ and $\text{C}\equiv\text{C}$, from the heats of combustion, or formation, of ethylene and acetylene respectively. Using the data for the C—H and C—C bonds together with the heat of dissociation of oxygen, the heats of combustion of alcohol, ether and acetone permit the energies of the C—O(H), C—O(C), and C=O bonds, respectively, to be derived. The actual values obtained depend on the data employed, and so cannot be regarded as precise, especially in view of the uncertainty concerning the exact additivity of

bond energies and of the heat of atomization of carbon; the results in Table 94 may be taken as reasonably self-consistent, although they are

TABLE 94. BOND ENERGIES

Bond	Energy	Bond	Energy	Bond	Energy
C—C	59 kcal.	C—O	70 kcal.	C—N	49 kcal.
C=C	100	C=O	150	C=N	94
C≡C	123			C≡N	150
		C—H	87		

only approximate. A comparison of the bond energies with the force constants in Table 92 shows a distinct parallelism, as is to be expected; the bond energy increases with the bond multiplicity, and this fact allows important conclusions to be drawn.²⁴

Bond Energies and Resonance.—The heat of formation of carbon monoxide from solid carbon and gaseous oxygen is 26.8 kcal., and so the bond energy is 210 kcal.; this is so much larger than the normal C=O value as to show clearly the presence of a triple linkage. Taking the heat of formation of carbon dioxide as 94.0 kcal., the energy of each carbon-oxygen bond is 168 kcal.; a comparison of this with the value for C=O in Table 94 indicates the possibility of resonance, with a triple bonded structure making an appreciable contribution. If the structure of carbon dioxide were O=C=O, the heat of formation from gaseous atomic carbon and oxygen should be about $2 \times 150 = 300$ kcal., assuming each bond to have the energy of a normal C=O linkage. The heat of formation of $+O \equiv C - O^-$ should involve that of a single and a triple carbon-oxygen bond; the former is 70 kcal., and the latter may be assumed equal at least to that in carbon monoxide, i.e., 210 kcal., so that the total is at least 280 kcal. In view of the fair proximity of the two energy quantities, it is evident that resonance should occur (cf. p. 112). Since the actual energy of formation of carbon dioxide is $2 \times 168 = 336$ kcal., the stabilizing energy due to resonance is about $336 - 300 = 36$ kcal. per mole. That is to say, as a consequence of resonance an additional 36 kcal. are required to dissociate 1 mole of carbon dioxide into its constituent atoms.

The existence of resonance in benzene has been demonstrated in an analogous manner. The heat of formation from gaseous atoms is 1039 kcal., but the sum of the energies of six C—H, three C—C and three C=C linkages is only 999 kcal., indicating a resonance energy of 40 kcal.* It is of interest to compare the observed and calculated heats of formation of cyclohexane; the former is 1398, and the sum of the energies of twelve C—H and six C—C linkages is also 1398 kcal. The exact agreement is partly fortuitous, but it is evident that the assumption of the additivity of bond energies is reasonably justified, and that the

* This value is not as accurate as the one derived below from heats of hydrogenation.

large discrepancy noted for benzene must have a definite significance. The comparison of observed and calculated bond energies was used by L. Pauling and J. Sherman (1933) to evaluate the resonance energies in a number of organic compounds.

Heats of Hydrogenation.—The accurate measurements of G. B. Kistiakowsky, *et al.* (1935 *et seq.*), mentioned on p. 203, have provided a valuable method for evaluating resonance energies in certain cases. The heat of addition of three molecules of hydrogen to the three double bonds of benzene is 49.8 kcal., whereas the heat of hydrogenation of the apparently analogous compound cyclohexene, containing one double bond, is 28.6 kcal. If there were no resonance the hydrogenation of benzene should be accompanied by $3 \times 28.6 = 85.8$ kcal.; the difference between this and the observed value, i.e., $85.8 - 49.8 = 36.0$ kcal. is the stabilization energy resulting from resonance. By comparing the heat evolved when $\Delta^{1:3}$ -butadiene is hydrogenated, i.e., 57.1 kcal., with twice the heat of hydrogenation of Δ^1 -butene, i.e., 2×30.3 kcal., the resonance energy of the former is seen to be 3.5 kcal., which is attributed to its conjugated system $-C=C-C=C-$. Other unsaturated compounds containing the same grouping have stabilization energies of the order of 2 to 4 kcal. per mole. If the two double bonds are separated by two or more single linkages, as in $\Delta^{1:4}$ -pentadiene and in $\Delta^{1:6}$ -hexadiene the results indicate the absence of resonance; since the double bonds are no longer conjugated, resonance does not occur.

Electronegativity Scale.—If the electron pair were shared equally between two atoms A and B, that is to say, if the bond were purely covalent in character, it would be expected that the energy of the A—B bond would be the mean of the values for the A—A and B—B bonds. In actual fact, the observed A—B bond energy is greater than the calculated (geometric) mean, and the square root of this difference is a measure of the relative electronegativities, i.e., the electron attracting tendencies (p. 108), of the atoms A and B. By assigning an arbitrary value to one particular element, e.g., hydrogen, it has been possible to construct an electronegativity scale (L. Pauling, 1932) for a number of elements. Some of the results obtained in this manner are as follows:

H	C	N	O	F	Cl	Br	I
2.1	2.5	3.0	3.5	4.0	3.0	2.8	2.5

The difference in the negativities of any two elements is a measure of the extent of the ionic character, and of the dipole moment, of the bond.²⁵

ELECTRON DIFFRACTION BY GASES AND VAPORS

Scattering of X-Rays and Electrons.—When X-rays are passed through a gas a definite pattern is obtained, consisting of a number of concentric rings due to diffraction by the atoms within the molecule (cf. p. 505), the separate molecules being too far apart for them to inter-

ferre with each other. The theory of the subject was worked out by P. Debye (1915) and P. Ehrenfest (1915), but it was not till 1929 that Debye applied his scattering equation to specific cases in order to obtain information concerning molecular dimensions. Since data of the same type can be obtained more conveniently by the diffraction of electrons, the X-ray method as applied to gases has received little further attention. The scattering of electrons is of two types, viz., coherent and incoherent. The latter is a result of the change in equivalent wave length of the electrons due to the Compton effect (p. 34); as it is only noticeable at small scattering angles, and is not important for the present purpose, it will be neglected. The coherent, or elastic, scattering is represented by an equation derived by H. Mark and R. Wierl (1930), similar to that applicable to X-rays; thus, when θ is the angle between the primary and the scattered electron beam, the intensity I of the latter is

$$I = k \sum_i \sum_j \psi_i \psi_j \frac{\sin x_{ij}}{x_{ij}}, \quad (107)$$

where k is a constant, ψ is the scattering factor of each particular atom for electrons, and x_{ij} is defined in terms of l_{ij} , the distance between the centers of the atoms designated i and j , viz.,

$$x_{ij} = 4\pi l_{ij} \frac{\sin \theta/2}{\lambda}, \quad (108)$$

where λ is the equivalent wave length of the electrons (p. 16). The value of ψ depends on the atomic number of the atom and its scattering factor for X-rays (p. 358), as well as on the angle θ and the wave length λ , but for appreciable scattering angles it may be replaced by the atomic number Z , so that (107) becomes

$$I/k = \sum_i \sum_j Z_i Z_j \frac{\sin x_{ij}}{x_{ij}}. \quad (109)$$

This equation means that the scattering intensity at any particular angle θ is proportional to the sum of all the $Z_i Z_j (\sin x_{ij})/x_{ij}$ terms for every possible pair of atoms in the molecule, the scattering due to individual atoms being included. Since the scattering depends on a quantity $(\sin x)/x$, and x involves θ , it is evident that I must pass through a series of maxima and minima as the scattering angle θ increases. It is for this reason that a fine beam of electrons after passage through a gas or vapor produces a series of concentric rings on a photographic plate. The dark rings represent positions of maximum scattering, whereas the lighter portions in between correspond to angles for which the electron scattering is a minimum.

Experimental Method.—A fine beam of electrons, accelerated by a potential of about 50,000 volts, meets at right angles a stream of the gas or vapor under

examination; the pressure within the apparatus is kept as low as 10^{-6} mm. by means of a pump. The beam is allowed to fall on a photographic plate for a short time, e.g., 0.5 to 5 sec., and the plate is then removed and developed; the positions of apparent maxima and minima of scattering intensity are marked by holding the plate up to a strong light. Photometric methods have been employed, but for various reasons they need not be discussed. The electron-accelerating potential is measured either by a suitable voltmeter, or by a milliammeter and a resistance; the equivalent wave length can then be calculated.

Scattering Intensity Curves.—The use of (109) can best be illustrated by an actual example, and carbon tetrachloride is a convenient case; the molecule may be regarded as tetrahedral, so that all four chlorine atoms are equivalent. The scattering equation is then

$$I/k = Z_C^2 + 4Z_{Cl}^2 + 12Z_{Cl}^2 \frac{\sin x_{Cl-Cl}}{x_{Cl-Cl}} + 8Z_C Z_{Cl} \frac{\sin x_{C-Cl}}{x_{C-Cl}}, \quad (110)$$

where the first term, on the right-hand side, is the scattering due to the carbon atom alone, and the second term is that produced by the four individual chlorine atoms. In these two quantities the corresponding value of l_{ij} is zero, and so by (108) x_{ij} is zero and $(\sin x_{ij})/x_{ij}$ must be unity. The third term represents the scattering by each pair of chlorine atoms, and since there are twelve pairs possible, this number precedes the scattering expression. The fourth term is for scattering between a carbon and a chlorine atom, eight such individual terms being involved. Since a tetrahedral configuration has been assumed, the distance l_{C-Cl} between the carbon and a chlorine atom is $\sqrt{3/8}$ times l_{Cl-Cl} , that is, between two chlorine atoms; hence x_{C-Cl} is equal to $\sqrt{3/8} x_{Cl-Cl}$, and (110) may be written

$$I/k = 6^2 + 4 \times 17^2 + 12 \times 17^2 (\sin x)/x + 8 \times 6 \times 17 (\sin \sqrt{3/8} x)/\sqrt{3/8} x, \quad (111)$$

where 6 and 17 are the atomic numbers of carbon and chlorine, respectively, and x is the form of (108), in which l is the chlorine-chlorine distance. It is now necessary to find the positions of maximum and minimum intensities, and in order to do this I/k is evaluated from (111) by taking various arbitrary numerical values of x from 0 to 50, the results being plotted on a curve. Since the actual intensities of scattering are not required, but merely the values of x for which maxima and minima occur, the first two terms on the right hand side of (111) may be neglected, as they make a constant contribution at all scattering angles. This is true in general, so that the scattering due to individual atoms need not be included in the intensity equation. Proceeding in the manner described it is found that the first three scattering maxima for carbon tetrachloride, according to (111), should be at values of x equal to 7.75, 13.9 and 20.6, respectively.

From the positions of the diffraction rings on the photographic plate, relative to that of the incident beam, the scattering angles corresponding to maximum intensities may be readily evaluated. Using the known wave length of the electron beam, determined from the accelerating potential, the values of $(\sin \theta/2)/\lambda$ for the first three maxima given by carbon tetrachloride vapor were found to be 0.217, 0.380 and 0.555×10^8 , respectively. It follows from (108), therefore, that

$$\begin{array}{ll} \text{(i)} & 7.75 = 0.217 \times 10^8 \times 4\pi l; \quad \therefore l = 2.81 \times 10^{-8} \text{ cm.}; \\ \text{(ii)} & 13.9 = 0.380 \times 10^8 \times 4\pi l; \quad \therefore l = 2.91 \times 10^{-8}; \\ \text{(iii)} & 20.6 = 0.555 \times 10^8 \times 4\pi l; \quad \therefore l = 2.95 \times 10^{-8}. \end{array}$$

A comparison of the positions of maximum scattering determined experimentally with those evaluated from (111) thus permits the calculation of the chlorine-chlorine distance in carbon tetrachloride as about 2.85 Å. It follows, therefore, that the carbon-chlorine distance is $\sqrt{3/8} \times 2.85$, i.e., 1.74 Å. By considering further maxima and minima of scattering the result obtained, which is generally accepted as reasonably accurate, is 1.76 Å.

Two or More Parameters.—In the example considered the configuration of the molecule can be expressed in terms of one parameter, viz., the chlorine-chlorine distance; whenever this is possible an exactly analogous method may be employed to evaluate interatomic dimensions in the molecule. In many cases, however, it is necessary to introduce two or more parameters; a molecule such as sulfur dioxide or chlorine monoxide, which is angular, requires two parameters, viz., an interatomic distance and the bond angle. It is now necessary to calculate the scattering curve for each of a number of different values of one of the parameters; for example, with chlorine monoxide L. E. Sutton and L. O. Brockway (1935) assumed bond angles of 60°, 90°, 120°, 150° and 180°, and derived a theoretical scattering intensity curve for each different angle. The experimental positions of the diffraction maxima and minima showed that the curve for 120° was most nearly correct, and the construction of further scattering curves in this vicinity gave $111 \pm 2^\circ$ as the bond angle. From the positions of the maxima and minima the chlorine-oxygen distance was found to be 1.71 ± 0.02 Å. If more than two parameters are required to describe the molecule then the number of trial intensity curves to be calculated is greatly increased, but the principle of the method whereby the correct values are eventually determined is the same as that explained above.

The Radial Distribution Method.—A form of the procedure used to interpret the X-ray diffraction patterns of liquids (p. 506) has been described by L. Pauling and L. O. Brockway (1935). The distribution function, which gives the product of the scattering powers of all volume elements constituting the molecule at every distance l apart, is calculated from the intensities of diffraction rings, and the values are plotted against l ; the maxima in the curve correspond

to distances separating important scattering centers, and hence to distances between atoms. The radial distribution method thus gives the principal interatomic distances without the necessity of assuming any particular molecular configuration; it suffers from the disadvantage, however, that if there are several maxima it may not be possible to identify them with particular atomic separations.

Applications of Electron Diffraction.—Electron diffraction studies have been used, as seen above, for two main purposes: the determination of molecular configurations and of interatomic distances. The oxygen bond angle in chlorine monoxide was shown to be about 111° and similar methods have been applied to determine the shapes of other simple molecules, e.g., dimethyl ether, dioxane and oxygen fluoride. The Cl—C—Cl angles in chloroform and methylene chloride were found to be $111 \pm 2^\circ$, proving the larger values suggested from dipole moments (p. 555) to be incorrect. The diffraction patterns also allow of a clear distinction between *cis*- and *trans*-modifications of dichloro- and dibromoethylene; in the *cis*-compounds the most important scattering atoms, i.e., the halogens, are much closer than in the *trans*-derivatives. The configurations of the triazo-group ($-\text{N}_3$) and the diazo-group in aliphatic compounds have been discussed from many points of view; according to parachor measurements ring structures should be applicable in each case, but the results of electron scattering have definitely excluded these configurations. There now appears to be little doubt that both methyl azide and diazomethane are linear molecules; the actual structures involve resonance among several forms of similar energy.

The interatomic distances determined by electron diffraction provide support for the rule of the additivity of atomic radii or, more correctly, of bond lengths (p. 382); when a discrepancy is observed it is necessary to find an explanation, and in many instances there is little doubt that resonance between two or more different structures is responsible. In carbon dioxide, for example, the carbon-oxygen distance is 1.13 Å., as compared with the additive values of 1.24 Å. for a double bond and 1.11 Å. for a triple bond; it is clear that the ionic states $^-\text{O}-\text{C}\equiv\text{O}^+$ and $^+\text{O}=\text{C}-\text{O}^-$ contribute very appreciably to the structure (see p. 112). The interatomic distances in carbon suboxide, cyanogen and nitrous oxide show the existence of resonance in each case. In benzene the carbon-carbon bond length is 1.39 Å., which may be compared with 1.54 Å. for a single link and 1.33 Å. for a double bond. Wave mechanical calculations have shown that the two Kekulé structures are the most important resonating states (p. 112), and so it may be supposed that each carbon-carbon bond in benzene partakes of double bond character to the extent of one-half. In graphite, where the distance between adjacent carbon atoms is 1.42 Å., the double bond contribution is only one-third, and it is of interest that if the four carbon-carbon distances, for single bond, double bond, and in benzene and graphite, are plotted against the corresponding amount of double bond character, a smooth curve is obtained. With the

aid of this curve the double bond contribution to a carbon-carbon linkage may be estimated from the bond length (L. Pauling and L. O. Brockway, 1937).²⁶

OPTICAL ROTATION

Optical Activity.—When a substance is capable of rotating the plane of polarization of light, it is said to have **optical rotatory power** and to be **optically active**. Compounds of this type may be divided into two main classes: one in which the optical rotation is evident in the crystal only, e.g., quartz and sodium chlorate, and the other wherein optical activity may be exhibited by the solid, liquid, in the pure state or as a solution, and gaseous forms. In the former category the ability to rotate the plane of polarization of light is related to the fact that the atoms in the crystal are arranged in the form of either a right-handed or a left-handed spiral (cf. p. 390); since this structure disappears on melting, the liquid shows no optical activity. Substances of the second type are active, however, by virtue of the asymmetry of the molecule itself, and the activity is retained in all physical states and in solution. A consideration of the nature and stereochemistry of compounds possessing optical activity is outside the scope of this book, but certain aspects of the subject are of physicochemical interest; these will be discussed in an elementary manner only, as many of the problems are very complex.

Specific and Molecular Rotation.—The optical rotatory power of a pure substance, particularly in the liquid state, is generally expressed in terms of its **specific rotation** or **specific rotatory power**, $[\alpha]$, defined by the equation

$$[\alpha] = \alpha/l\rho, \quad (112)$$

where α is the measured angle of rotation for a column of liquid l decimeters in length, and ρ is the density of the substance. The specific rotation of a solution is given by

$$[\alpha] = \alpha/l\rho f, \quad (113)$$

f being the *weight* fraction of the active substance in the solution of density ρ ; the relationship is sometimes written in the form

$$[\alpha] = 100\alpha/l\rho w = 100\alpha/lc, \quad (114)$$

where w is the weight of substance in 100 g. and c in 100 cc. of solution. The **molar rotation**, or **molar rotatory power**, $[M]$, is obtained when the specific rotation $[\alpha]$ is multiplied by the molecular weight M of the optically active compound, and the result divided by 100; thus,

$$[M] = M[\alpha]/100. \quad (115)$$

The angle through which the plane of polarization of light is rotated by a substance depends to some extent on the temperature, and particularly on the wave length of the light; both these quantities should therefore be stated, e.g., $[\alpha]_D$, indicates a measurement for the D -line of sodium. For solutions the results are also dependent on the nature of the solvent and on the concentration; the effect of the latter is negligible for dilute solutions.

Linear, Circular and Elliptical Polarization.—If a beam of polarized light proceeding in a straight line has all its vibrations in one plane, it is said to be **linearly polarized**; two such beams with equal amplitudes, but with their planes of polarization at right angles, may be combined to give three different types of vibration. The results are best shown graphically, as in Fig. 130, where AB

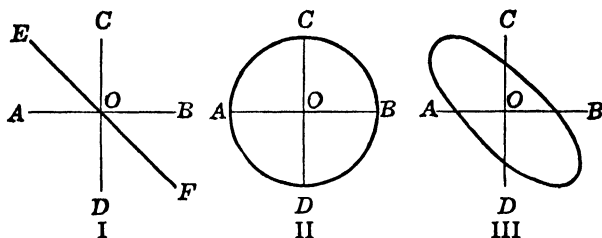


FIG. 130. Linear, circular and elliptical polarization

and CD represent twice the amplitude of the two linear vibrations at right angles about the center of vibration O . If the vibrations are so timed as to commence at the same instant at A and C , respectively, and to reverse simultaneously at B and D , then the combination gives a linear vibration EF (Fig. 130, I); the resultant is, therefore, still linearly polarized. In general when the difference of phase of the two components is either zero or an even number of quarter wave lengths, the resultant is a linearly polarized ray. If the phase difference is an odd number of quarter wave lengths, as occurs when one vibration, e.g., AB , is passing through O just as the vibration at right angles, viz., CD , is starting from C , the resultant of the two linear vibrations is a circular motion, as represented in Fig. 130, II; a suitable combination of two linearly polarized rays of the same amplitude therefore gives a **circularly polarized** ray. If the component AB is moving to the right through O when CD is starting down from C , the circular vibration is clockwise, but if the component AB should be moving to the left at that instant, the resultant vibration will be anti-clockwise. It is apparent, therefore, that two circularly polarized rays of the same amplitude, but with vibrations in opposite senses, will combine to give a linearly polarized ray in the direction CD , since the two AB components cancel each other. When the phase difference between two linear vibrations at right angles is neither zero nor exactly an odd or even number of quarter wave lengths, the combination of the linear vibration gives an elliptical vibration; for the case depicted in Fig. 130, III one vibration is an eighth part of a wave length behind the other. A resultant ray of this type is said to be **elliptically polarized**.

Optical Rotation and Circular Polarization.—A ray of linearly polarized light may be regarded as equivalent to two circularly polarized beams of equal amplitude but with opposite senses of rotation; the right- and left-hand circular vibrations meet in a line which determines the position of the plane of polarization. In the passage through an inactive material the velocities of transmission of both components are affected to the same extent, and the position of the line in which they meet, and hence the plane of polarization, remains unchanged. In an optically active medium, however, it may be supposed that the two circularly polarized rays have different velocities in the direction of the beam of light; the result is that the components now meet on a line which is inclined to the original, and hence the plane of polarization is rotated through a definite angle. Since the refractive index of a given medium is related to the velocity of light in it, it appears that optical rotation may be regarded as due to a difference in the refraction of right- and left-circularly polarized light. This interpretation of optical activity was proposed by A. J. Fresnel (1825), who showed that the angle of rotation α per cm. of medium was given by

$$\alpha = \frac{\pi}{\lambda} (n_l - n_r), \quad (116)$$

where λ is the wave length of the light, and n_l and n_r are the refractive indices for left- and right-circularly polarized light, respectively. If n_l exceeds n_r the angle of rotation is in one direction, but if n_r is greater than n_l it will be in the opposite sense. In view of the relatively small wave length of visible light the difference between the two refractive indices need only be very small, e.g., 10^{-6} to 10^{-6} , to give appreciable angles of rotation.

Rotatory Dispersion.—Since the refractive index of light depends on the wave length (p. 533), it is reasonable to expect that the rotatory power of a substance will change in a somewhat analogous manner; the variation of angle of rotation with the wave length of light is known as the **rotatory dispersion**. As a result of his not too accurate observations on quartz, J. B. Biot (1817) suggested the inverse square relationship $\alpha = A/\lambda^2$, between the angle of rotation α and the wave length λ of the light, A being a constant; this was known to be approximate, and an improved equation, involving an additional constant,

$$\alpha = A + B/\lambda^2, \quad (117)$$

which is analogous to the simple Cauchy formula (p. 533) for refractive dispersion of light, was used by P. von Lang (1863) and by J. Stephan (1864). An important contribution to the subject of rotatory dispersion was made by P. Drude (1900), who deduced theoretically, by methods now said to be partially unsound, an equation that undoubtedly represents the actual behavior of the great majority of optically active substances. In its simplest form the **Drude equation**, which is somewhat similar to the

Sellmeier dispersion equation (p. 533), may be written

$$[\alpha] = \frac{k}{\lambda^2 - \lambda_0^2}, \quad (118)$$

λ_0 and k being characteristic of the given substance; λ_0^2 is called the **dispersion constant** and k the **rotation constant** (T. M. Lowry, 1914). According to theoretical considerations the wave length λ_0 is that at which the molecule absorbs radiation strongly, and this should correspond to the frequency of the equivalent classical electronic oscillator in the molecule (p. 585). Hence the value of λ_0 should approximate, as it does, to the wave length of an absorption band in the ultraviolet or visible portions of the spectrum.

Simple and Complex Dispersion.—Within and close to an absorption band the Drude equation, like the refractive dispersion equation, does not apply. As the experimental wave length λ approaches λ_0 the value of α should become very large, but this is not found in practice, although it might be ideally if the absorption occurred at a single line,* instead of a band. As will be seen shortly, in the vicinity of an absorption band new phenomena appear, but in regions of complete transparency an equation involving one or more Drude terms represents in a very satisfactory manner the variation of angle of rotation with wave length of the polarized light. When a one-term equation is sufficient, as is frequently the case, the substance is said to exhibit **simple rotatory dispersion**, but if two or more terms are required the dispersion is said to be **complex** (T. M. Lowry, 1914). An illustration of simple rotatory dispersion is provided by *sec.*-octyl alcohol (methylhexylcarbinol); the observed specific rotations at 20° c. are expressible over a considerable range of wave length by the single term Drude equation †

$$[\alpha] = \frac{3.176}{\lambda^2 - 0.0244}.$$

On the other hand, ethyl tartrate at 20° exhibits complex dispersion and its optical activity is represented by the two-term equation

$$[\alpha] = \frac{25.005}{\lambda^2 - 0.03} - \frac{20.678}{\lambda^2 - 0.056}.$$

The remarkable accuracy with which these equations reproduce the experimental results will be evident from Table 95.

If the single term Drude equation is obeyed the plot of $1/[\alpha]$ against λ^2 should be a straight line, but the reverse of this statement, namely, that a straight line plot indicates simple dispersion, is not necessarily true,

* R. W. Wood (1905) studied the magnetic rotatory dispersion (see p. 608) of sodium vapor, and found the rotation to increase markedly as the wave length of the light used approached that of either of the two *D*-lines.

† In writing the Drude equation it is the convention to express λ and λ_0 in microns (μ), i.e., 10^{-4} cm. units; thus 5000 Å., or 5000×10^{-8} cm., is 0.5μ , and λ would be 0.5.

TABLE 95. OPTICAL ROTATION AND THE DRUDE EQUATION

Sec. Octyl Alcohol			Ethyl Tartrate	
λ	$[\alpha]$ obs.	$[\alpha]$ calc.	$[\alpha]$ obs.	$[\alpha]$ calc.
6708 Å.	7.46°	7.47°	6.87°	6.87°
6362	8.35	8.35	7.43	7.43
5893	9.85	9.85	7.82	7.82
5218	12.82	12.81	7.61	7.61
4678	16.34	16.34	5.45	5.46
4271	20.10	20.10	0.7	0.5
4046	22.80	22.80	-5.1*	-5.1
3860	25.44	25.50	-12.2	-12.2

* Interpolated.

for the graphical method serves only to show up appreciable departures from simple behavior. Where the discrepancies are small they only become evident if algebraic analysis is employed, and this procedure is always recommended. A number of cases have been found of substances whose behavior is apparently satisfied by a single term equation, but which are believed, for theoretical reasons, e.g., the position of more than one active absorption frequency, to require at least two terms. In such instances the additional terms may be small, and only appreciable at very short wave lengths, or they may cancel each other almost exactly. It is probable, theoretically, that rotatory dispersion should always be represented by a number of terms, one for each electronic oscillator affected by the asymmetry of the molecule. The terms found necessary, within the limits of experimental error, probably represent, at least approximately, the main contributions to the total activity.²⁷

Normal and Anomalous Dispersion.—In general, a two term Drude equation may be written in the form

$$[\alpha] = \frac{k_1}{\lambda^2 - \lambda_1^2} + \frac{k_2}{\lambda^2 - \lambda_2^2}. \quad (119)$$

If the two terms have the same sign, e.g., both positive or both negative, or if the terms have different signs with $k_1 > k_2$ and $\lambda_1 > \lambda_2$, so that the first term is always larger than the second, then $[\alpha]$ increases or decreases regularly with decreasing wave length; the dispersion is then said to be **normal**. There is, however, the possibility that while $k_1 > k_2$, the constant λ_1 is less than λ_2 ; then if the two terms are of opposite sign, the specific rotation will first increase, reach a maximum and then decrease and eventually the sign of the rotation will become reversed. A substance exhibiting this type of behavior, known as **anomalous dispersion**, is ethyl tartrate, as the results in Table 95 show. As the wave length decreases the specific rotation rises to a maximum at 5893 Å.; it then decreases and is zero at 4240 Å., after which the rotation, now of opposite sign, increases numerically. The reversal of the rotation occurs when the two terms of the Drude equation, being opposite in sign,

become equal; that is when

$$\frac{\lambda^2 - \lambda_1^2}{\lambda^2 - \lambda_2^2} = \frac{k_1}{k_2}. \quad (120)$$

The difference between simple, normal complex and anomalous complex rotatory dispersion is shown diagrammatically in Fig. 131. It must be

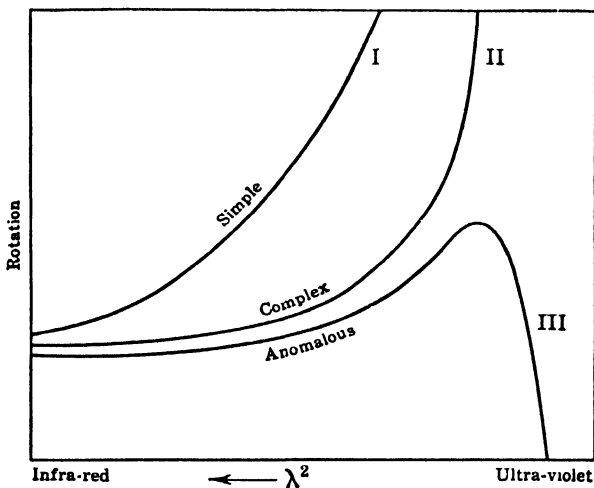


FIG. 131. Normal and anomalous rotatory dispersion

emphasized, for reasons which will appear shortly, that the reversal of sign accompanying anomalous dispersion occurs outside, and generally at some distance from, an absorption band.

Position of Absorption Bands.—Many saturated aliphatic compounds, not possessing a strongly chromophoric group, exhibit simple dispersion and the value of λ_0^2 lies between 0.020 and 0.025, so that λ_0 is approximately 1500 Å. The result is in agreement with the accepted view (p. 582) that substances of this type have absorption bands in the far ultra-violet. The presence of an unsaturated, or any other recognized chromophoric, grouping produces a considerable change in λ_0 . For example, in the series of aliphatic secondary alcohols of general formula $\text{CH}_3\text{CH}(\text{OH})\text{R}$, the value of λ_0 is approximately constant throughout at 0.024, but replacement of R by a phenyl group changes λ_0 to 0.043; this implies, as might be expected, an absorption band nearer to the visible region. A few cases may be quoted of substances whose absorption bands come within the measurable part of the spectrum; the observed values are compared, in Table 96, with those calculated from the optical rotation by means of the Drude equation. The difference between the two sets of results is probably due to the fact noted above that the one- or two-term Drude equation can only be an approximation to the truth. As a general rule the λ_0 value for a given substance is almost the same in the pure

TABLE 96. POSITION OF ABSORPTION MAXIMUM

Compound	λ_0 from Drude Equation	Absorption Maximum
Camphorquinone	4730 Å.	4650 Å.
Camphor	2950	2880
γ -Nonyl nitrite	3680	~3700
Ethyl tartrate	2370	~2300

liquid as in nonpolar solvents; a similar conclusion has been recorded with reference to the positions of observed absorption bands (p. 586).

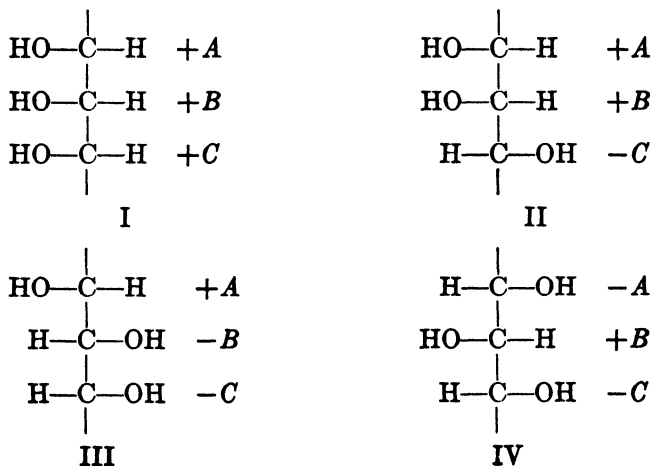
Chromophoric Centers and Optical Rotation.—The representation of rotatory dispersion by a series of Drude terms is of importance since it provides a means of assessing partial rotations, that is, the contributions to the observed rotation of the various light absorbing, i.e., chromophoric, groups, or more strictly of the electrons capable of undergoing energy transitions of relatively small magnitude, which are affected by the asymmetric field of the molecule. The introduction of a chromophoric group into an optically active molecule does not necessarily imply that the absorption band will make a contribution to the total rotation (L. A. Tschugaeff, 1913). Nicotine has an absorption maximum at 2650 Å., and the corresponding chromophore is active, as is evident from the fact that λ_0 in the Drude equation is equivalent to 2550 Å.; the introduction of a $>C=O$ group, in methyl nicotine, gives an absorption band at 2950 Å., which is expected for a ketonic group, but the rotatory dispersion remains simple and λ_0 is equivalent to 2668 Å. The new chromophore thus makes no contribution to the optical activity (T. M. Lowry and H. K. Gore, 1931). It is only when the electronic oscillator is able to couple with the "active" electrons in the remainder of the molecule that the additional group provides a term in the dispersion equation. The possibility of coupling depends, in the first instance, on the proximity of the chromophoric group to the asymmetric center, and the inducing action of the latter on the former has been called the **vicinal function** or **vicinal action** (W. Kuhn, 1930). It should be emphasized that mere proximity is not the only factor determining this vicinal action, for the nature of the groups present is important. A striking illustration is provided by the compounds of the type $XCH_2CH(CH_3)Pr$, where X is bromine or iodine. If Pr is the *isopropyl* group then dispersion is normal, the absorption band of the halogen providing the main rotation, but if it is a *n*-propyl group, the dispersion is anomalous and the contribution of the halogen band is reversed. The influence of the position of the chromophoric group is also often remarkable. The introduction of a CH_2 group between the active carbon atom and the aldehyde group in $CH_3CH(C_2H_5)CHO$ changes the molar rotation for the *D*-line from $+20.3^\circ$ to -8.7° , but a second CH_2 group increases the value to $+12.0^\circ$ (P. A. Levene, 1936). Many observations of this kind, which are difficult to explain, are to be found in the literature of optical activity.

In some instances it appears that a chromophoric group is able to dominate the optical activity of the whole molecule; this occurs, for example, in camphor, when the absorption of the $>C=O$ group provides the main partial rotation. It was suggested (T. M. Lowry, 1924) that the coupling of a strong chromophoric group to an asymmetric complex might result in the production of induced asymmetry, later called **induced dissymmetry**, in the former. In general the new center of activity produced in this way would have a rotation opposite to that of the remainder of the molecule, so that anomalous dispersion might

result, as with camphor. The rotatory dispersion of the secondary alcohols $\text{RCH}(\text{OH})\text{R}'$ is simple, but that of their esters, in certain solvents, is anomalous; this has been attributed to the dissymmetry induced in the $-\text{C}\overset{\text{O}}{\parallel}-$ group, which thus becomes optically active. In view of the fact that the contribution is to be ascribed to the electronic oscillator, rather than to the group as a whole, it would appear preferable to refer to the phenomenon as **induced anisotropy** (W. Kuhn, 1930) of the substituent or of its absorption band. This expression does not imply any stereochemical changes, such as are suggested by the use of the terms asymmetry or dissymmetry.

In general, anomalous dispersion may be attributed to the presence of two chromophoric centers in the same molecule which contribute partial rotations of opposite sign, or to the existence of two molecular species in equilibrium, each having one chromophore. The fact that tartaric acid and ethyl tartrate exhibit anomalous dispersion, whereas the methylene and dimethylene tartaric acids are normal, has presented a problem of great interest; it has been suggested that tartaric acid and its ethyl ester exist in two forms resulting from two possibilities of hydrogen bond formation between the hydrogen of the hydroxyl group and the ketonic oxygen of the carboxyl or ester group (T. M. Lowry, 1923).²⁸

Optical Rotation and Chemical Structure.—In spite of considerable experimental investigation concerning the relationship between the specific or molar rotation and chemical structure, very few general rules have been discovered. The principle of **optical superposition**, first proposed by J. H. van't Hoff (1894), applies particularly to isomeric substances possessing several asymmetric carbon atoms. The rotatory power of such a substance is, according to the principle, the *algebraic* sum of the contributions of each separate carbon atom, which is a definite amount and is independent of the configurations of the other atoms. The four pentose sugars each have three asymmetric carbon atoms, the configurations of which are as follows:



If the contributions of the three carbon atoms in I (ribose) are assumed to be $+A$, $+B$ and $+C$, respectively, the values for the others are as shown, the reversal of the configuration of an atom resulting in reversal of sign. According to the superposition principle, therefore, the specific rotation of II (arabinose) should be the sum of the values for the other three isomeric sugars; thus,

$$+A + B - C = (+A + B + C) + (+A - B - C) + (-A + B - C). \quad (121)$$

The principle is approximately applicable to the four isomeric menthylamines, each of which has three asymmetric carbons (J. Read, 1930), although it is apparently necessary to assume a reversal of sign of the contribution of one of the atoms resulting from a change in the configuration of the other two.

The **isorotation rules**, proposed by C. S. Hudson (1909), represent an advance on van't Hoff's principle of superposition, particularly in regard to their wider applicability. Hudson suggested (i) that the rotation of the first carbon atom in many carbohydrates was only slightly changed by alterations in the structure of the remainder of the molecule; and (ii) that changes in the structure of the first carbon atom had relatively little effect on the rotation of the molecule. The first rule is supported by the fact that reversal of the configuration of the $-\text{CH}(\text{OCH}_3)$ group in a series of methyl glucosides produces almost the same change in the molar rotation in a pentose, a hexose and a biose; thus,

$$\begin{array}{lll} \alpha\text{-methyl-}d\text{-xyloside} & - \beta\text{-methyl-}d\text{-xyloside} & = 359^\circ, \\ \alpha\text{-methyl-}d\text{-glucoside} & - \beta\text{-methyl-}d\text{-glucoside} & = 375^\circ, \\ \alpha\text{-methyl-}d\text{-gentibioside} & - \beta\text{-methyl-}d\text{-gentibioside} & = 361^\circ. \end{array}$$

The contribution of the terminal carbon atom thus appears to be independent of the remainder of the molecule. The second rule may be verified by comparing the α - and β -forms of different derivatives of the same sugar; the configurations of the terminal carbon atoms are thus changed, but the remainder of the molecule is unaltered. If the contribution of the end group is assumed to be positive in the α -form and negative in the β -form, the sum of the two molecular rotations should give the contribution of the remainder of the sugar, and this should be constant irrespective of the nature of the end group, in agreement with the following results.

$$\begin{array}{lll} \alpha\text{-glucose} & + \beta\text{-glucose} & = 237.2^\circ, \\ \alpha\text{-methyl glucoside} & + \beta\text{-methyl glucoside} & = 242.0^\circ, \\ \alpha\text{-glycol glucoside} & + \beta\text{-glycol glucoside} & = 235.1^\circ. \end{array}$$

Hudson's rules have been applied in various ways to throw light on the structure of sugars.²⁹

A generalization of a different type has appeared from the work of M. Betti (1907-20) and of H. G. Rule (1929-31) on the changes of rotation resulting from the replacement of a hydrogen atom by a series of different

groups. In many cases the influence of a given substituent is related to its dipole moment and to its effect on the strength of acetic and benzoic acids. These results suggest that the polarity of the substituent, or perhaps of the molecule as a whole, is an important factor in determining the magnitude of optical rotation. Modern theories associate optical activity with the motion of certain electrons in an asymmetric electric field (p. 607), and it is reasonable to expect changes of rotation to be related to alterations in the distribution of electrical charges in the molecule.

Solvent Effects.—The influence of the solvent on optical activity is very complex, but one important generalization, which has a theoretical basis, has been discovered. If there is no interaction between the optically active solute and the solvent, the effect of the latter can be eliminated by dividing the observed specific (or molar) rotation by $(n^2 + 2)/3$, where n is the refractive index of the solvent. The resulting values are almost constant, and are equal to the rotation in the vapor state, when the refractive index is unity. For solutions of a polar solute in polar solvents, the observed rotatory power of a given substance decreases, in general, with increasing dipole moment in a series of related solvents. In nonpolar solvents the specific rotation sometimes decreases with increasing concentration; this is presumably due to the increasing polarity of the medium.³⁰

Circular Dichroism: The Cotton Effect.—In the course of an investigation of the optical activity of potassium chromium tartrate solutions *within* the absorption band, A. Cotton (1896) discovered some interesting facts to which are applied the general name of the **Cotton effect**. In the first place, the optical rotation within the band was found to be anomalous; it increased to a maximum near the absorption band, then decreased to zero within the band, and finally increased in the other sense to a second maximum, as shown in Fig. 132. It was also observed that right- and left-circularly polarized light are absorbed to a different extent by an optically active substance, so that a linearly polarized ray of wave length within an absorption band is converted into an elliptically polarized one on passage through the material. This phenomenon is known as **circular dichroism**; its magnitude is expressed by the ellipticity ϕ , in radians, of the emergent beam per cm. of medium defined by the equation

$$\phi = \frac{\pi}{\lambda} (\kappa_l - \kappa_r), \quad (122)$$

where κ_l and κ_r are the absorption indices * for left- and right-circularly polarized light. The ellipticity increases numerically as the absorption band is approached, reaches a maximum and then decreases (Fig. 132); the wave length at the maximum corresponds approximately to that of the center of the band, and also to that at which the rotation changes sign. The sign of the ellipticity depends on the relative values of κ_l and κ_r , and L. Natanson (1908) proposed the following rule relating the sign to that of the rotation: the sign of the elliptical vibration is the same as the sign of the rotation on the long wave length side of an optically active absorption band, whereas on the short wave length side the signs are reversed.

* The absorption index κ is defined by Beer's law (or Lambert's law), p. 581, in the form $I = I_0 e^{-4\pi\kappa d/\lambda}$, where λ is the wave length of the absorbed light.

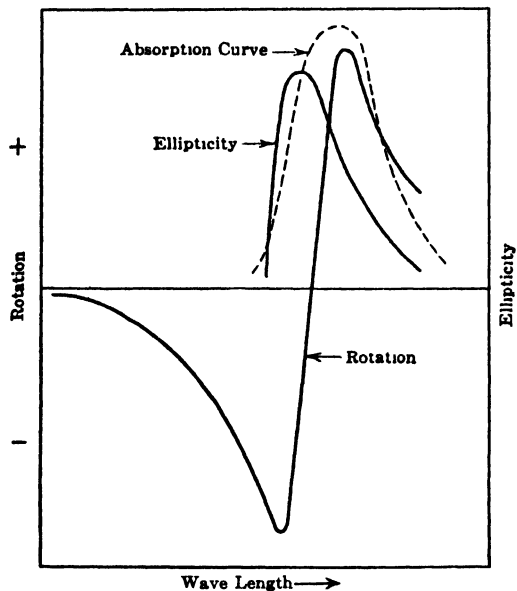


FIG. 132. The Cotton effect

The circular dichroism is conveniently expressed by means of the quantity g , viz.,

$$g = (\kappa_l - \kappa_r) / \kappa, \quad (123)$$

where κ is the absorption index for ordinary light at the same wave length. This quantity has been called the **anisotropy factor** (R. Kuhn, 1930) or the **dissymmetry factor** (T. M. Lowry, 1933). Its value lies primarily in the theoretical deduction that it should be directly proportional to the wave length.

The Cotton effect has been confirmed for various colored compounds with absorption bands in the visible region, as well as for colorless substances possessing bands in the near ultraviolet.* In fact measurements of circular dichroism are now regarded as an essential part of a complete study of the optical properties of a compound, and the results are desirable for an adequate test of theories of optical activity. The observations are, of course, possible only when the extinction coefficients in the absorption band are not too large, and sufficient light is transmitted to permit the necessary readings to be taken.

Strictly speaking, the angle of rotation becomes zero and is reversed in sign in an absorption band only when the electron or electronic system producing that band is the sole, or by far the most important, contributor to the total optical activity. When this is not the case, the dispersion curve shows the same type of inflection as in Fig. 132, but the rotation does not pass through zero in the middle; this is the result of a steadily increasing, or decreasing, contribution due to another part of the molecule added to that of the absorption band under consideration.

* An active substance exhibits the Cotton effect only in its own absorption band, and, further, only if the electronic system responsible for the band also contributes to the optical activity. There is no Cotton effect in the absorption band of an added inactive material.

The phenomena associated with the Cotton effect are of great theoretical significance, and equations have been deduced relating the specific rotation and circular dichroism within an absorption band to the wave length. By combination of these equations with empirical expressions for the variation of absorption intensity in the band, calculations have been made of the partial contributions of the "active" groups in compounds, such as methyl- α -azidopropionate, $N_3CH(CH_3)COOCH_3$, where part of the rotation is produced by the electrons of the $-N_3$ group, which has an absorption band with a maximum at about 2900 \AA ., and part by the remainder of the molecule, chiefly the $-COOCH_3$ group which absorbs at shorter wave lengths. The continuity of the partial rotation of the latter, without any change of direction, in the absorption region of the other active group, may be used as a test for the reliability of the equations employed.²¹

Experimental Methods.—The rotation of the plane of polarization of light is measured with some form of polarimeter; monochromatic light of known wave length first passes through a polarizer, e.g., a Nicol prism, then through the liquid under examination, and finally through another prism to act as analyzer. The relative positions of the two prisms for the maximum transmission of light are observed with and without the active substance; the difference in angle gives the required rotation. Various devices are employed to enable the correct positions of the polarizer and analyzer to be determined with accuracy; these are described in books on experimental physical chemistry. For measurements in the ultraviolet region photographic methods of recording are employed; the use of glass must, of course, be avoided. Circular dichroism was first studied by Cotton by observing the absorptive power of the medium for right- and left-circularly polarized light separately, but more recently direct methods for the determination of the ellipticity have been devised.

Theories of Optical Activity.—The first important attempt to provide a theoretical basis of optical activity, was made by P. Drude (1896) who assumed the presence in an active molecule of electric charges vibrating in spirals; this assumption led him to the development of the rotatory dispersion equation (p. 598), but it has now been shown that the mathematical treatment does not correspond to the proposed model and, moreover, the model would not be optically active. Modern work in this field may be said to commence with M. Born's theory of coupled vibrators (1915) which represents a molecule by a system of electrical particles which become polarized, in the electrical sense (p. 534), and oscillate in the electric field of a light wave; the coupling of at least two such oscillators would give rise to optical activity. The principle of coupled vibrators was applied to a definite molecular model, in which the vibrators are at right angles, by W. Kuhn (1930); although it has been successful in the analysis of the Cotton effect, the general theory is of limited value since it is not applicable to molecules with an asymmetric carbon atom, but only to certain optically active *spiro*-compounds containing two rings at right angles. M. Born (1934) has made a further extension of the theory based on the simple tetrahedral model of an asymmetric molecule; he considered four coupled oscillators, and derived an equation for the rotatory power in terms of the characteristic wave lengths of the oscillators and of their strengths, as well as of directly measurable quantities. Another type of theory has been discussed by R. de Malleman (1925, 1930) and by S. F. Boys (1934); it involves a model in which four different groups are attached to a central carbon atom, and the

rotations are calculated in terms of the electron polarizabilities, and hence the refraction equivalents, of the groups. Other theories of optical activity have been proposed by J. G. Kirkwood (1937), and by E. U. Condon, W. Altar and H. Eyring (1937); the latter authors have shown that contrary to the views of Born, Kuhn and others, dynamic coupling between two or more electron oscillators is not essential to optical activity, for the effects can be produced by a single chromophoric electron moving in the unsymmetrical force field of the asymmetric molecule. The mathematical treatment of theories of optical activity is too difficult for further discussion here.³²

MAGNETIC ROTATORY POWER

The Faraday Effect.—The phenomenon of the rotation of the plane of polarization of light by transparent substances when placed in a magnetic field was discovered by M. Faraday (1846), and hence is often called the **Faraday effect**. The light must travel along the lines of the field and the sense of the rotation depends on the direction of the magnetic field. Magnetic rotation differs in an important respect from natural optical rotation. If the beam of light, in the latter case, is reflected back along its course, so that it passes once in each direction through the active medium, the resultant rotation will be zero. Since magnetic rotation depends on the direction of the field, and this is the same for the light passing in both directions, the effect is doubled. It is evident from (116) that the magnetic field must alter the refractive indices of right- and left-circularly polarized light to different extents; this may be due to relatively different changes in the characteristic absorption frequencies for the two rays.

Investigations of the Faraday effect by E. Verdet (1854–58), and others, showed the angle of rotation (α) to depend on the nature of the substance; it is proportional to the length (l) of the column of the latter which the light traverses and to the strength (H) of the magnetic field; thus

$$\alpha = \omega l H, \quad (124)$$

where ω is the Verdet constant for the experimental material. This constant is the angle of rotation in a field of 1 gauss for a 1 cm. depth of liquid; it is almost independent of temperature, but varies markedly with the wave length of the light. The Verdet constant for water at 20° c. is 0.01308 min. of arc for the sodium *D*-line; this datum is useful as it can be utilized to calculate the constants for other substances by comparing their rotation with that of water in the same magnetic field.

Molar Magnetic Rotation.—An extended study of the magnetic rotation of organic compounds, and its relation to structure, was made by W. H. Perkin (1884–1907); he used a powerful electromagnet, the pole pieces containing holes through which passed the polarimeter tube. A sodium flame was employed as the source of light. Other workers have obtained the necessary magnetic field by winding coils of wire round a tube containing the liquid and passing a powerful current. For the

comparison of magnetic rotatory powers of different substances, Perkin used water as a standard and defined the **molar magnetic rotation** $[M]$ by the expression

$$[M] = M\alpha\rho'/M'\alpha'\rho, \quad (125)$$

where M is the molecular weight, α the angle of rotation and ρ the density of the given liquid, and M' , α' and ρ' are the same quantities for water; the angles of rotation α and α' are measured in the same tube and in the same magnetic field. In any homologous series the addition of a CH_2 group results in a mean increase of 1.023 in the magnetic rotation, so that the value of $[M]$ for any compound in a series can be expressed as

$$[M] = 1.023n + s, \quad (126)$$

where n is the number of CH_2 groups and s is a constant for the given series. For normal paraffins s is 0.508, for primary alcohols 0.699, for aldehydes 0.261, for ketones 0.375, and for acids 0.393; thus for the primary alcohol $\text{C}_n\text{H}_{2n+1}\text{OH}$,

$$[M] = 1.023n + 0.699,$$

so that the series constant is the magnetic rotation equivalent of a hydrogen atom and the characteristic group of the series, i.e., the $-\text{OH}$ group. From the series constants, starting with that of the paraffins, which gives the value for hydrogen, it is possible to derive rotation equivalents for various atoms. Perkin's results are given in Table 97; they are the mean of values which often differ appreciably. Conjugation of double bonds is accompanied by magnetorotatory exaltation which persists even in

TABLE 97. MAGNETIC ROTATION EQUIVALENTS

Hydrogen	0.254	Chlorine	1.734
Carbon	0.515	Bromine	3.562
Oxygen (in $-\text{OH}$)	0.191	Iodine	7.757
Oxygen (in $-\text{CHO}$)	0.261	Nitrogen (in $-\text{NH}_2$)	0.483
Oxygen (in $>\text{CO}$)	0.375	Double bond	1.11

closed ring structures; for example, the magnetic rotation of benzene should be the sum of 6×0.515 , for 6C , 6×0.254 for 6H and 3×1.11 for three double bonds, that is 7.94, compared with the observed value 11.28. The discrepancy, 3.34, is somewhat smaller than that between observed and calculated magnetic rotations of hexatriene, $\text{GH}_2=\text{CH}-\text{CH}=\text{CH}-\text{CH}=\text{CH}_2$, viz., 3.85, which also contains three conjugated double bonds. Magnetic rotation results have been applied in connection with problems of molecular structure, particularly in the terpene series, and also to determine approximately the condition of substances capable of undergoing keto-enolic tautomerism.

MAGNETIC PROPERTIES

Paramagnetism and Diamagnetism.—All substances and all forms of matter possess magnetic properties in some degree (M. Faraday, 1845),

and three different types of behavior have been recognized: they are referred to as ferromagnetism, paramagnetism and diamagnetism. If two magnetic poles m_1 and m_2 are separated by a distance r in any medium, the force between the poles is F , where

$$F = \frac{1}{\mu} \cdot \frac{m_1 m_2}{r^2}. \quad (127)$$

The quantity μ ,* assumed by convention to be equal to unity in vacuo and is approximately unity in air, is the **magnetic permeability** of the medium; it is a measure of the tendency of the magnetic lines of force to pass through the medium in comparison with a vacuum or air. For **paramagnetic** substances μ is slightly greater than unity and for **ferromagnetic** materials it is of the order of 10^3 or more. Magnetic lines of force, therefore, prefer to pass through a paramagnetic, and especially a ferromagnetic, substance rather than through a vacuum. On the other hand, when the permeability is less than unity the material is **diamagnetic** and lines of force prefer to travel through an evacuated space. A consequence of this difference of permeability is that when placed in a magnetic field a paramagnetic bar will tend to set itself with its length parallel to the field, whereas if diamagnetic it will set itself at right angles; in the former instance the lines of force travel the maximum, and in the latter case the minimum, distance through the bar.

If H is the strength of a magnetic field the product μH is called the magnetic induction B in a given medium, and the intensity \dagger of magnetism I induced by the field in that medium, is determined by the equation $B = 4\pi I + H$; hence,

$$\mu H = 4\pi I + H \quad \text{and} \quad \mu = 4\pi I/H + 1. \quad (128)$$

The relationship between H and the intensity I of induced magnetization is defined by $I = \kappa H$, where κ is the **volume susceptibility** of the medium; combining this with (128) gives

$$\mu = 1 + 4\pi\kappa. \quad (129)$$

For most practical purposes the **mass susceptibility**, or **specific susceptibility**, χ , obtained by dividing κ by the density ρ of the material, thus $\chi = \kappa/\rho$, is used; hence,

$$\mu = 1 + 4\pi\rho\chi. \quad (130)$$

Since μ is greater than unity for ferromagnetic and paramagnetic materials, χ is obviously positive, its value being of the order of 10^3 for the former, but only about 10^{-3} c.g.s. unit for the latter. For diamagnetic

* This should not be confused with dipole moment, for which the same symbol is employed.

† This is the magnetic moment for unit volume, in the direction of the field.

substances, however, μ is less than unity and χ must be negative; the susceptibilities have been found to be approximately -10^{-6} c.g.s. unit.

The properties of diamagnetic materials are in general independent of temperature and field strength, but for paramagnetic compounds the susceptibility is nearly inversely proportional to the absolute temperature (P. Curie, 1895). Ferromagnetic substances are abnormal in their behavior; the susceptibility decreases with increasing temperature, and above a certain critical temperature, the Curie point, they become paramagnetic. The value of χ varies with the field strength, and the phenomena of hysteresis and remanescence, familiar properties of iron, are observed. The chief ferromagnetic substances are iron, cobalt and their alloys; their magnetic properties are more of physical than of chemical interest and so will not be considered further. Para- and diamagnetism have special chemical applications which merit discussion.

It has been seen (p. 65) that every rotating or spinning electron is associated with a magnetic moment; if the quantum numbers of the electrons in an atom or molecule are such that the moments cancel each other, the substance is diamagnetic. Since the great majority of molecules contain an even number of electrons, mostly with completed electron groups, their resultant magnetic moment is zero and diamagnetism is a common occurrence. Paramagnetic behavior is observed when one or more of the subgroups of electrons is incomplete, and hence it is to be expected in molecules and ions with an odd number of electrons (G. N. Lewis, 1923). It should be noted that for elements, particularly metals, it is necessary to distinguish between the magnetic properties of the single atoms, i.e., in the gaseous state, and of the material in bulk. By the Stern-Gerlach method (p. 64) it is found that silver and copper atoms have a resultant magnetic moment, but in the massive state they are diamagnetic. The electrons in metals are known, however, to behave in an exceptional manner (p. 379).

The magnetic properties of a substance in the elementary state are often totally different from those in the ionic or combined condition: for example, ferrous and ferric ions, and complex ions with iron as the central atoms, are not ferromagnetic. Further, metallic copper is diamagnetic and so also are the cuprous salts, but the cupric compounds are paramagnetic; the differences in these instances are accounted for by changes in the electron groups (cf. p. 88).

Determination of Susceptibility.—The general method for determining magnetic susceptibility is based on that of G. Gouy (1889). A thin cylinder, or tube, containing the material under examination, is suspended vertically between the poles of an electromagnet, the axis of the cylinder being at right angles to the lines of force. At its lower end the tube must be in a region of uniform magnetic field, but near the upper portion the field strength should be negligible. The downward magnetic force F acting on the material is

$$F = \frac{1}{2}(\kappa_1 - \kappa_2)aH^2, \quad (131)$$

where κ_1 and κ_2 are the volume susceptibilities of the material and air (0.03×10^{-6}), respectively, a is the area of cross section of the specimen and H is the field strength. The value of F is determined by attaching the tube containing the material to the arm of a balance, and finding the weight necessary to compensate for the pull * on it when the magnetic field is switched on. For purposes of absolute measurement the field strength H would have to be determined but this can be avoided if the apparatus is calibrated with a substance, e.g., ferrous ammonium sulfate, which has an accurately known susceptibility, or with water.

For the study of liquids a form of G. Quincke's method (1885) may be used. The substance is contained in a U-shaped tube, one limb being wide and the other narrow. The latter is placed between the poles of an electromagnet with the surface of the liquid in the uniform portion of the field; the magnet is switched on and the change in level h is observed, then

$$h g \rho = \frac{1}{2}(\kappa_1 - \kappa_2) H^2, \quad (132)$$

κ_1 and κ_2 being the susceptibilities of the liquid and the gas above it.

To obtain the specific susceptibility the value of κ is divided by the density of the substance. In chemical work, the atomic (χ_A) or molar (χ_M) susceptibilities are generally employed; these are the result of multiplying the specific susceptibility by the atomic or molecular weight. Provided there is no interaction, and no considerable departure from ideal behavior, the susceptibility of a mixture of two substances is in many cases the sum of the separate contributions of the constituents. Thus if χ is the measured specific susceptibility of a mixture containing *weight* fractions f_1 and f_2 of two components of susceptibilities χ_1 and χ_2 , then

$$\chi = f_1 \chi_1 + f_2 \chi_2. \quad (133)$$

This equation can be used to determine the susceptibility of a substance from measurements in solution; it has been applied to the study of salts in water, the susceptibility of the latter being known accurately.

Diamagnetic Substances.—It was shown by P. Langevin (1905) that when a magnetic field is applied to a system of electrons moving about a relatively heavy nucleus, the velocities of the electrons are altered with the result that there is a precession (cf. p. 57) of the electron orbits. The angular velocity of precession is numerically equal to $eH/2mc$, where H is the field strength, e and m are the charge and mass of the electron, and c is the velocity of light, and this produces a change in the magnetic moment of the system of $-e^2 H r_i^2 / 4mc^2$ for each electron, r_i^2 being the mean square of the radius of the electron orbit as projected on a plane at right angles to the direction of the field. For an atom containing n electrons, the resultant change of moment ΔM_μ is

$$\Delta M_\mu = - \frac{e^2 H}{4mc^2} \sum_n \overline{r_i^2}. \quad (134)$$

If the atom possesses spherical symmetry the $\sum \overline{r_i^2}$ term can be replaced

* This will be negative for a diamagnetic substance.

by $\frac{2}{3}\sum \bar{r}^2$, where \bar{r}^2 is the mean square of the *actual* orbital radius, so that

$$\Delta M_{\mu} = -\frac{e^2 H}{6mc^2} \sum_n \bar{r}^2, \quad (135)$$

the $\sum \bar{r}^2$ term implying summation of \bar{r}^2 for all the n electrons. The change in moment given by (135) applies to all substances, but for those which are paramagnetic and possess a large magnetic moment it is negligible. When the electrons are in completed subgroups, however, and their resultant moment is zero, it is this change of moment on application of a magnetic field which is the origin of diamagnetism. To determine the atomic susceptibility it is necessary to multiply ΔM_{μ} by the Avogadro number N and to divide by H ; * thus,

$$\chi_A = -\frac{Ne^2}{6mc^2} \sum_n \bar{r}^2. \quad (136)$$

Inserting the known values for N , e (in c.m. units), m and c , then

$$\chi_A = -2.84 \times 10^{10} \sum_n \bar{r}^2. \quad (137)$$

Since the radius of an atom is about 10^{-8} cm., \bar{r}^2 is about 10^{-16} and hence χ_A should be of the order of 10^{-6} e.m. (c.g.s.) unit; this is in agreement with observation for diamagnetic substances.

The equation for χ_A given above was derived from the classical theory of the electron; it has been improved by the methods of quantum mechanics, and better agreement is thus obtained between calculated and experimental susceptibilities especially for small atoms. By making certain simplifying assumptions the new equation reduces to the same form as (136), and so the latter may be used as a first approximation. It should be noted that the equation was deduced for a system involving one nucleus only, and so it should hold for monatomic molecules and ions; it appears also to be fairly satisfactory for other molecules possessing spherical symmetry.

Results: Elements and Compounds.—Nonmetallic elements, with the exception of oxygen and S_2 , are diamagnetic and so also are a number of metals. It may be taken, as a general rule, that the inert gases and the elements of the B subgroups are diamagnetic. Many of such elements behave abnormally, however, as their susceptibilities change with temperature. In any group of the periodic table the atomic susceptibility increases with increasing atomic weight, that is, with increasing size; this is in general agreement with (136), for the larger the atom the larger the $\sum \bar{r}^2$ term. The molar susceptibilities of simple molecules are also approximately related to their dimensions; the value for hydrogen, -3.9×10^{-6} , is low, and the following results have been obtained for other familiar gases:

* By the definitions given above $\chi_A = AI/H\rho$, where A is the atomic weight; the magnetic moment per unit volume, which is equal to I , is $\Delta M_{\mu} \times N\rho/A$ and so $\chi_A = \Delta M_{\mu} \times N/H$, assuming the substance to be monatomic.

N ₂	N ₂ O	CO ₂	NH ₃	SO ₂	HCl
-12	-19	-19	-19	-18	-25 × 10 ⁻⁶ c.g.s. unit.

The result for chlorine, -40×10^{-6} , is unexpectedly high. There is generally little difference between the atomic or molar susceptibilities in liquid and gaseous states.

Ions.—The molar susceptibilities of a number of salts have been measured in aqueous solution, and since constant differences are observed when the cation is changed but the anion remains the same, and *vice versa*, it is probable that the ionic susceptibilities are additive. By assuming the ratio of the values for two ions, e.g., Cs⁺ and Cl⁻, to be the same in the solid salt as in solution, it has been found possible to assign values to a number of ions; some of the results are quoted in Table 98. They are in

TABLE 98. SUSCEPTIBILITIES PER GRAM ION OF IONS IN SOLUTION *

F ⁻	-10.3 × 10 ⁻⁶	Na ⁺	-5.6 × 10 ⁻⁶		
Cl ⁻	-25.1	K ⁺	-14.0	Ca ⁺⁺	-5.5 × 10 ⁻⁶
Br ⁻	-36.4	Rb ⁺	-23.0	Sr ⁺⁺	-14.8
I ⁻	-54.2	Cs ⁺	-37.3	Ba ⁺⁺	-27.8

* The values recorded must be regarded as approximate only.

harmony with the values calculated by methods involving wave mechanics. It is of interest that the ionic susceptibilities are just what would be expected from the ionic radii; in each vertical series the value increases as the ions become larger. When ions with the same inert gas structure are compared, e.g., Cl⁻, K⁺ and Ca⁺⁺, the susceptibilities are seen to decrease as the charge changes from -1 to +2; the dimensions of the ions decrease in the same order (cf. p. 383). The atomic susceptibilities of neon, -7.6×10^{-6} , and of argon, -19×10^{-6} , fit into their correct places in the first two horizontal series. The susceptibilities of ions in the solid state are generally close to those in solution, except for small and bivalent ions when the values in solution are lower, probably because of interaction with water molecules.²³

Susceptibility Equivalents.—A very complete study of the diamagnetism of organic and inorganic compounds was made by P. Pascal (1908-13) who found the molar susceptibility to be an additive and constitutive property. By application of the methods employed for other additive properties, it is possible to ascribe susceptibility contributions to various atoms and to structural factors, starting with the fact that the increase of χ_M in a number of homologous series is close to -11.7×10^{-6} for each additional CH₂ group. There is lack of agreement among the results in the literature, but those recorded in Table 99 are probably reasonably correct. The susceptibility equivalents of carbon, nitrogen, sulfur and the halogens are almost the same as for the pure elements, and so they are true atomic susceptibilities. A number of structural factors appear to be necessary, and the values given in Table 99 include those most commonly encountered. When allowing -1.5 for a benzene

TABLE 99. ATOMIC AND STRUCTURAL SUSCEPTIBILITY EQUIVALENTS

Hydrogen	-2.85×10^{-6}	Fluorine	-11.5×10^{-6}
Carbon	-6.0	Chlorine	-20.1
Oxygen		Bromine	-30.6
-O- in ethers and alcohols	-4.6	Iodine	-44.6
O= in ketones	+1.73	Sulfur	-15.0
O= in esters and acids	-3.4	Nitrogen in amines	- 5.6
		Benzene ring	- 1.5
Double bond between:		Triple bond between:	
Carbon atoms	+5.47	Carbon atoms	+ 0.77
Nitrogen atoms	+1.8	Carbon and nitrogen	- 0.8
Nitrogen and carbon	+8.2		

nucleus it is unnecessary to add the equivalents of the double bonds; thus the sum of the contributions of six carbon and six hydrogen atoms is -53.1 and together with -1.5 the total is -54.6 , compared with the experimental value, -55.0×10^{-6} . The suggestion has been made that the positive equivalent of a double bond is to be ascribed to the drawing together of the nuclei and hence of the electron orbits, with a consequent decrease in the Σr^2 term, or its equivalent; this cannot be a complete explanation, for the influence of a triple bond should then be greater, whereas it is less.

As an illustration of the additivity of diamagnetic susceptibilities the results for methyl acetate may be considered. The experimental value of χ_M is -43.6×10^{-6} , whereas the sum of the atomic equivalents is -43.1×10^{-6} , made up of -18.0 for three carbon atoms, -17.1 for six hydrogen atoms, and -4.6 and -3.4×10^{-6} for the two oxygen atoms. In view of the marked structural influences it is possible that measurements of diamagnetism could be used to decide between alternative structures; one interesting application is to follow the course of polymerization processes, since these invariably involve a change in the type of linkage. The susceptibility constants in Table 99 are used when it is required to calculate the susceptibility of an atom in the combined state. The molar susceptibility of cadmium oxalate is -53.1×10^{-6} , while the contribution of the C_2O_4 group, assuming the same type of structure as in an ester, would be -28.0×10^{-6} , leaving -25.1×10^{-6} as the value of χ_A for cadmium. A result of the same order has been obtained for a large number of cadmium compounds, including the so-called "sub-chloride," Cd_4Cl_7 , indicating that the latter is a bivalent compound of cadmium probably with some free metal in solid solution (J. F. Spencer, 1935).³⁴

Paramagnetism.—A simple theory of paramagnetism was developed by P. Langevin (1905) who regarded each paramagnetic molecule, i.e., each molecule with an unbalanced electron orbit, as behaving like a small magnet with a definite moment. The axis of each molecular magnet will tend to arrange itself so as to be parallel to the direction of an applied field, but this tendency is opposed by the thermal motion of the molecules.

By the use of the Maxwell-Boltzmann distribution law, as described in connection with dipole moments (p. 544), Langevin showed that the mean moment \bar{m} of a *gaseous* molecule in the direction of the external magnetic field is given by

$$\bar{m} = \frac{m_{\mu}^2}{3kT} H, \quad (138)$$

where m_{μ} is the actual magnetic moment of each molecule and H the field strength, provided the latter is not too large. By the definitions given above χ_M is equal to $MI/\rho H$, where M is the molecular weight, and since M/ρ is the molar volume V , then $\chi_M = VI/H$. Since I is equal to the magnetic moment, in the direction of the applied field, per unit volume, the product VI is the total magnetic moment per mole. This can also be obtained by multiplying the average moment \bar{m} per molecule by the Avogadro number; hence VI is equal to $\bar{m}N$, and $\chi_M = \bar{m}N/H$. Substitution in (138) then gives the atomic or molar *paramagnetic* susceptibility χ_{μ} ; * thus,

$$\chi_{\mu} = \frac{Nm_{\mu}^2}{3kT} \quad \text{or} \quad m_{\mu} = \sqrt{\frac{3kT\chi_{\mu}}{N}}. \quad (139)$$

The effective moment M_{μ} per mole is the product of the moment m_{μ} per single molecule and the Avogadro number, so that

$$M_{\mu} = \sqrt{3kNT\chi_{\mu}} = \sqrt{3RT\chi_{\mu}}. \quad (140)$$

If R is expressed in c.g.s. units, then division by 5585 (see p. 65) gives the value in terms of molar Bohr magnetons; † thus,

$$M_{\mu} = \sqrt{3RT\chi_{\mu}}/5585 = 2.83 \sqrt{\chi_{\mu}T} \text{ Bohr magnetons}. \quad (141)$$

According to Curie's law (p. 611) the susceptibility of most paramagnetic substances is inversely proportional to the temperature, so that $\chi_{\mu} = C_M/T$, where C_M is a constant for each substance; hence (141) is sometimes written in the form

$$M_{\mu} = 2.83 \sqrt{C_M} \text{ Bohr magnetons}. \quad (142)$$

The determination of the effective moment of a paramagnetic substance involves the measurement of χ_{μ} ; the actual susceptibility, obtained as described on p. 612, gives the sum of the paramagnetic and diamagnetic susceptibilities, but as the latter is small in comparison with the former it can be neglected. It is preferable, however, to make an allowance for the diamagnetic contribution, and this can usually be done by means of the atomic susceptibilities, such as those in Table 99, or by making measurements with a closely related diamagnetic compound.

* The symbol χ_{μ} is used instead of χ_M to indicate the paramagnetic, as distinct from the diamagnetic, susceptibility.

† It is sometimes the practice to express magnetic moments in terms of the empirical Weiss magneton, at one time believed to be a fundamental unit of magnetism (P. Weiss, 1911); the Weiss magneton is 4.96 Bohr magnetons.

Paramagnetism and Electron Spin.—It has been recorded (cf. p. 65) that, apart from certain terms that can be neglected in all but a few instances, the magnetic moment of an atom or ion is expressed, in molecular Bohr magnetons, by the formula

$$M_{\mu} = \sqrt{J(J+1)}g \quad (143)$$

[cf. equation (84), p. 65], which reproduces in a satisfactory manner the observed values for the ions of the rare earth elements. When the same equation is applied to the ions of the transitional elements, however, it leads to results in disagreement with experiment; the explanation of this discrepancy is that in the liquid, i.e., solution, and solid states, in which the great majority of paramagnetic substances are studied, the perturbing effect of the surrounding molecules destroys the contribution of the orbital momentum (L) to the magnetic moment. The latter is thus determined entirely by the spin quantum number (S), and hence by the spin moments of the unpaired electrons. In these circumstances J and S are identical, and (143) becomes

$$M_{\mu} = 2 \sqrt{S(S+1)} \text{ Bohr magnetons,} \quad (144)$$

where 2 is the corresponding value of the Landé splitting factor g (p. 65). The magnetic moments of the rare-earth ions are produced by electrons in the incompleated 4f subgroup (p. 91), which is protected from the perturbing influence of external molecules and ions by the complete outer shells of 5s and 5p electrons. For these ions it is necessary, therefore, to use the original equation (143), but for other paramagnetic molecules and simple ions the incomplete electron subgroup is the outermost one, and the postulated effect of surrounding molecules, in the crystal or in solution, can be readily understood.

Since each unpaired electron contributes $\frac{1}{2}$ towards the value of S , the latter is equal to half the number of such electrons in the molecule or ion; it is, therefore, possible to obtain the results in Table 100, relating

TABLE 100. MAGNETIC MOMENT AND UNPAIRED ELECTRON SPINS

No. of unpaired electrons	Total Spin Angular Momentum (S)	M_{μ} Bohr magnetons
0	0	0
1	$\frac{1}{2}$	1.73
2	1	2.83
3	$1\frac{1}{2}$	3.88
4	2	4.90
5	$2\frac{1}{2}$	5.91

the expected moment of a paramagnetic substance to the number of electrons with unpaired spins. By putting M_{μ} in (141) equal to 1.73 molar Bohr magnetons, it is found that $\chi_{\mu}T$ should equal 0.374; hence at ordinary temperatures, about 290° K., the paramagnetic molar susceptibility χ_{μ} for a single unpaired electron should be about 1300×10^{-6} e.m.

unit. A value of this order is frequently encountered since most paramagnetic molecules have only one odd electron.⁵⁵

Elements.—The majority of metals, particularly those of the A subgroups, are paramagnetic, but their susceptibilities are generally much below that required for a single unpaired electron. Even many metals with an odd number of electrons, e.g., sodium, have low susceptibilities in the solid and liquid forms. The reason is that in the metallic state the electrons do not remain associated with the nuclei (p. 379) but form a separate mobile system, and W. Pauli (1927) has shown that such free electrons would be expected to give rise to a small paramagnetism independent of temperature; this has been found for many, but not all, metals. One of the most interesting cases of paramagnetism of elements is that of oxygen, both liquid and gaseous; the measured susceptibility corresponds to a M_μ value of 2.8 Bohr magnetons, indicating the presence of two electrons with uncoupled spins. The proposed electronic structure for molecular oxygen (p. 107) must therefore satisfy this requirement. It is of interest that the S_2 molecules in sulfur vapor are also paramagnetic, and so they must have an electronic structure similar to C_2 molecules.

Odd-Electron Molecules and Free Radicals.—G. N. Lewis (1923) suggested that compounds containing an odd number of electrons would prove to be paramagnetic, and the few gaseous compounds falling within this category, viz., nitric oxide, nitrogen dioxide and chlorine dioxide, do indeed exhibit paramagnetism. The molecular susceptibilities are all about 1300×10^{-6} e.m. unit at 20°C. , in accordance with the requirements for a single electron with uncoupled spin. Since free radicals, such as those of the triphenylmethyl type, contain an odd electron, they should exhibit appreciable paramagnetism; the first evidence of this nature was obtained by N. W. Taylor (1926), who found 140×10^{-6} for χ_μ of α -naphthylidiphenylmethyl, measured in a 7 per cent solution in benzene at 20°C. The theoretical result for a single electron should be 1300×10^{-6} e.m. unit, and so it appears that about 11 per cent of the material is present in the free radical form in the solution employed. Measurements of magnetic susceptibility of solids and solutions have been employed by Eugen Müller (1935) and by S. Sugden (1936), and others, to confirm the existence of free radicals and in some instances to determine the extent to which hexa-arylethanes, C_2R_6 , dissociate into free radicals, $\cdot CR_3$. Other substances, not free radicals in the narrow sense of the term, but possessing an odd number of electrons, e.g., $(p\text{-CH}_3\text{OC}_6\text{H}_4)_2\text{NO}$, $(\text{KSO}_3)_2\text{NO}$ and $\text{C}_6\text{H}_5\text{C(OK)C}_6\text{H}_5$, have the paramagnetic susceptibility required by theory for one uncoupled electron. Results of this type are of interest as they establish the fact that the metal ketyls, such as benzophenone potassium, $\text{C}_6\text{H}_5\text{C(OK)C}_6\text{H}_5$, do actually exist in the monomeric form (S. Sugden, 1934). On the other hand, it has been shown that hypophosphoric acid has the formula $\text{H}_4\text{P}_2\text{O}_6$, and not H_2PO_3 . If the latter were correct the sodium and silver salts, NaHPO_3 and Ag_2PO_3 , would contain an odd number of electrons

and should be paramagnetic; the fact that they are diamagnetic proves them to be $\text{Na}_2\text{H}_2\text{P}_2\text{O}_6$ and $\text{Ag}_4\text{P}_2\text{O}_6$, respectively. Similarly, the diamagnetism of mercurous chloride in solid, solution and vapor states shows that HgCl cannot be present in any phase, for this should be paramagnetic.³⁶

Attempts have been made to prove the existence of organic "bi-radicals," that is, of molecules having an even number of electrons two of which are uncoupled,* for such substances would have a magnetic moment of 2.83 Bohr magnetons. Several compounds which might be expected to behave as bi-radicals apparently isomerize to form quinonoid structures in which the electron spins are coupled. In a few cases, however, e.g., in porphyrindine derivatives in which no quinone form is possible, and in some biphenyl compounds, definite magnetic evidence for the presence of bi-radicals has been obtained (E. Müller, 1939).

Simple Ions.—Simple ions possessing an inert gas structure, e.g., of alkali and alkaline earth metals and of the halogens, or complex ions having an effective atomic number (p. 111) equal to the atomic number of an inert gas, and in fact all ions with complete electron subgroups, are diamagnetic. It is only among the rare earths and the transitional elements that paramagnetism appears. A study of this subject has raised a number of problems of special interest in connection with valence. In Table 101 are given some mean values for the experimental magnetic moments of cations, in Bohr magnetons. The results are little different in solution and in the solid state; in calculating the paramagnetic moment (M_μ) allowance has been made for the diamagnetic susceptibility of the anions. It will be at once evident from the table that the possession of an even number of electrons is no criterion of diamagnetism for ions. Several attempts have been made to interpret the results in Table 101;

TABLE 101. PARAMAGNETIC MOMENTS OF CATIONS

Ion	Number of Electrons	M_μ	Uncoupled Electrons
Cr^{+++}	21	3.85	3
$\text{Cr}^{++}, \text{Mn}^{+++}$	22	4.85	4
$\text{Mn}^{++}, \text{Fe}^{+++}$	23	5.8	5
$\text{Fe}^{++}, \text{Co}^{+++}$	24	5.3	4
Co^{++}	25	5.0*	3?
Ni^{++}	26	3.2	2
Cu^{++}	27	1.9	1
$\text{Cu}^+, \text{Zn}^{++}$	28	0	0

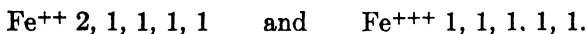
* Much lower values have been obtained with solid compounds.

the most successful, which accounts also for the magnetic properties of complex ions, is that of L. Pauling (1931), who considers that the paramagnetism of the transitional elements is determined by the number of uncoupled d electrons. This is supported by the fact that the largest number of unpaired electrons indicated by the results above is 5, and this

* The molecules O_2 and S_2 are simple examples of bi-radicals.

is also the number of d subgroups. Further, all ions with completed d levels are diamagnetic, e.g., Cu^+ , Zn^{++} , Ag^+ and Cd^{++} . The normal state of any ion is considered to be the one in which the maximum number of electrons, compatible with the Pauli principle, have unpaired spins (cf. Hund's rule, p. 67).

The electronic structure of the iron atom may be written $1s^2 2s^2 2p^6 3s^2 3p^6 3d^6 4s^2$, all the subgroups up to and including $3p^6$ being complete; the outermost groups of the ferrous and ferric ions must, therefore, contain six, i.e., $3d^6$, and five, i.e., $3d^5$, electrons, respectively, in d levels. Since there are only five d levels, each capable of containing a pair of electrons, the distribution of electrons among them, according to the postulate given above, must be

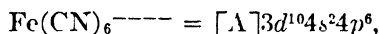


There are consequently four unpaired electron spins in the ferrous and five in the ferric ion, in accordance with the results in Table 101. The method can be applied to account for the magnetic properties of the other ions in the table, and the same principle is applicable to the corresponding ions of the next two series in the periodic system, except that $4d$ and $5d$ electrons, respectively, are involved.

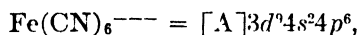
Complex Ions.—The empirical rule, proposed by L. O. Welo and O. Baudisch (1925), that the number of uncoupled electrons in a complex ion having a transitional metal as its central atom is equal to the difference between the nearest inert gas atomic number and the effective atomic number (p. 111) of the complex ion, is applicable in certain instances. It appears to hold, for example, for most of the six-covalent complexes formed by the elements from chromium to nickel, inclusive, but the rule breaks down when four-covalent ions, e.g., $\text{Ni}(\text{CN})_4^{++}$, are involved. The suggestions made by Pauling for simple ions can, however, be extended to account for the magnetic properties of many complexes. It is necessary to take into consideration the stereochemical configuration in each case, the basic assumption being that, in addition to d electrons, only the two s and the six p electrons of the next principal level, which correspond to the inert gas octet, are concerned in coordinate bond formation. If the coordination number is four, and the attached groups are arranged tetrahedrally, the s and p electrons only are involved, leaving the d electrons unaffected. If, on the other hand, the arrangement is planar, the d electrons are involved in bond formation and the magnetic properties are consequently altered. The $\text{Ni}(\text{CN})_4^{--}$ ion is planar, and consequently the vacant $3d$ levels take part in bond formation, giving the structure of the outermost groups $3d^{10} 4s^2 4p^4$, which represents a diamagnetic substance; the $3d$ level being filled by its ten electrons, there can be none with uncoupled spins. The glyoximes of nickel, which are four-covalent, are diamagnetic, and should consequently be planar; if this is the case unsymmetrical derivatives, e.g., the nickel derivative of benzylmethylglyoxime, should exist in *cis*- and *trans*-forms, as has been actually

found by S. Sugden (1932, 1935). In $\text{Ni}(\text{NH}_3)_4^{++}$, however, the NH_3 groups are arranged tetrahedrally, so that the d levels are not affected; the electronic structure is $3d^84s^24p^6$, and so the ion should be paramagnetic, since the eight electrons in the d level must have the distribution 2, 2, 2, 1, 1, among the five sublevels, leaving two unpaired electrons. The actual moment of 2.7 Bohr magnetons is in agreement with this deduction. In general, a tetrahedrally coordinated complex ion will have the same magnetic moment as the central ion, e.g., $\text{Fe}(\text{N}_2\text{H}_4)_2^{++}$, $\text{Co}(\text{N}_2\text{H}_4)_2^{++}$, $\text{Ni}(\text{NH}_3)_4^{++}$, and $\text{Cu}(\text{NH}_3)_4^{++}$, for which the values are close to those for Fe^{++} , Co^{++} , Ni^{++} , and Cu^{++} , respectively. Although nickel carbonyl $\text{Ni}(\text{CO})_4$ has a tetrahedral structure it must be diamagnetic; here four groups are coordinated to a nickel atom, and since eight electrons are added the outermost electron groups must be $3d^{10}4s^24p^6$. The structure corresponds to that of an inert gas, viz., krypton, in which all the sublevels are completed and so the compound must be diamagnetic.

In complexes of coordination number six, some of the d electrons must take part in bond formation and the magnetic moment of the ion depends on the resulting configuration. In the ferrocyanide ion, for example, twelve electrons, one pair from each cyanide ion, are added to the ferrous ion giving



where $[\text{A}]$ represents the argon structure (cf. p. 88). Since the $3d$ subgroup is obviously complete the ion should be, as it is, diamagnetic. The ferricyanide ion, however, will have the structure



so that there is one uncoupled $3d$ electron; the magnetic moment should then be 1.8 Bohr magnetons, in agreement with the experimental value of 2.0. The ions $\text{Mn}(\text{CN})_6^{---}$, involving trivalent manganese, and $\text{Cr}(\text{CN})_6^{---}$, with bivalent chromium, both have the outermost electronic grouping $3d^84s^24p^6$; there should thus be two unpaired $3d$ electrons and the magnetic moment should be 2.83 in each case, in agreement with the observed results of approximately 3.0 Bohr magnetons. In accordance with the above postulates the six-coordination compounds of trivalent cobalt ($\text{Co}^{+++} = 3d^6$), trivalent iridium ($\text{Ir}^{+++} = 5d^6$), and of quadrivalent platinum ($\text{Pt}^{++++} = 5d^6$) are all diamagnetic. It must be mentioned that complexes containing the NO group behave abnormally.

Pauling has suggested the possibility that in compounds involving strongly electronegative coordinated groups, e.g., fluorine, the latter may be joined to the central atom by bonds which are largely ionic in character; thus the FeF_6^{---} ion might be $[\text{Fe}^{++}(\text{F}^-)_6]$. Since the electron groups in the fluorine ion are all completely coupled, the magnetic moment should be the same as that of the ferric ion, i.e., 5.9 Bohr magnetons, as compared with 1.8 magnetons for the apparently similar $\text{Fe}(\text{CN})_6^{---}$ ion. The experimental evidence is in favor of this suggestion.

The theory of paramagnetism described above is of considerable interest, although there are some difficulties to which reference must be made in conclu-

sion. In the first place, it has been assumed that the ferric ion is Fe^{+++} , whereas it is probably $\text{Fe}(\text{H}_2\text{O})_6^{+++}$ both in the solid state and in solution. To account for the five uncoupled electrons it is necessary to assume, for this and other simple ions, that the molecules of water are held by electrostatic forces only. Further, in the complex ions considered above the d subgroup is often surrounded by completed s and p levels as in the rare-earth ions; equation (144), however, is apparently still applicable although, according to the views expressed on p. 617, it should not be under these conditions. It is suggested that the attachment of the coordinated groups introduces new perturbing effects.

The subject has been considered from a different point of view by J. H. Van Vleck (1935); according to his argument if the coordinated group is not held to the central atom by strong forces then the atom retains its spin multiplicity and its magnetic moment unchanged. This is assumed to happen in FeF_6^{---} where the forces are ionic, and presumably also in the hydrates. In the $\text{Fe}(\text{CN})_6^{---}$ ion, however, the strong binding upsets the Russell-Saunders coupling and gives the lowest possible, instead of the highest possible, number of uncoupled spins.

REFERENCES

- Bhatnagar, S. S. and Mathur, K. "Magnetochemistry."
 Brode, W. R. "Chemical Spectroscopy."
 Debye, P. "Polar Molecules."
 Fajans, K. "Radioelements and Isotopes: Chemical Forces and Optical Properties, etc."
 Herzberg, G. "Molecular Spectra and Molecular Structure."
 Klemm, W. "Magnetochemie."
 Kremann, R. "Physikalische Eigenschaften und Chemische Konstitution."
 Le Fèvre, R. W. J. "Dipole Moments."
 Lowry, T. M. "Optical Rotatory Power."
 Pauling, L. "The Nature of the Chemical Bond."
 Selwood, P. W. "Magnetochemistry."
 Sidgwick, N. V. "The Covalent Link."
 Smiles, S. "Chemical Constitution and Physical Properties."
 Smyth, C. P. "Dielectric Constant and Molecular Structure."
 Stuart, H. A. "Molekülstruktur."
 Sugden, S. "The Parachor and Valency."
 Sutherland, G. B. B. M. "Infra-Red and Raman Spectra."
 Van Vleck, J. H. "The Theory of Electric and Magnetic Susceptibilities."

1. Sugden, *J. Chem. Soc.*, 125, 32 (1924); *op. cit.*; Mumford and Phillips, *ibid.*, 155 (1928); 2112 (1929); Sidgwick, *et al.*, *ibid.*, 1876 (1930); Lindemann, *Ber.*, 63, 1650 (1930); Mann and Purdie, *J. Chem. Soc.*, 1549 (1935); 873, 1503 (1936); Anderson, *ibid.*, p. 1283; Gibling, *ibid.*, 299, 304 (1941); Hovorka, *et al.*, *J. Am. Chem. Soc.*, 62, 1096, 2372 (1940); 63, 1097 (1941); Quayle, *et al.*, *ibid.*, 66, 935, 938 (1944).
2. Lorentz, *Ann. Phys.*, 9, 641; Lorenz, *ibid.*, 11, 70 (1880); Bruhl, *Z. phys. Chem.*, 7, 1, 140 (1891); Eisenlohr, *ibid.*, 75, 585 (1910); "Spektrochemie organischer Verbindungen" (1912); Auwers and Eisenlohr, *Ber.*, 43, 806 (1910); Allsopp and Willis, *Proc. Roy. Soc.*, 153A, 379, 392 (1936); Stevels, *Trans. Farad. Soc.*, 33, 1381 (1937); Huggins, *J. Am. Chem. Soc.*, 63, 1916 (1941).
3. von Steiger, *Ber.*, 54, 1381 (1921); Fajans and Knorr, *ibid.*, 59, 249 (1926); Smyth, *Phil. Mag.*, 50, 361, 715 (1925); *op. cit.*; Samuel, *Z. Phys.*, 49, 95 (1928); 50, 380 (1929).
4. Fajans, *et al.*, *op. cit.*; *Z. Phys.*, 23, 1 (1924); *Z. Elek.*, 34, 502 (1928); *Z. phys. Chem.*, A147, 241 (1930); B24, 103 (1934); *J. Am. Chem. Soc.*, 64, 3023 (1942).
5. Debye, *Phys. Z.*, 13, 97 (1912); *op. cit.*

6. Smyth, *op. cit.*, Smyth, *et al.*, *J. Am. Chem. Soc.*, **61**, 3063 (1943); *J. Chem. Phys.*, **7**, 1085, (1939); **9**, 352 (1941).
7. Rodebush, *et al.*, *J. Chem. Phys.*, **8**, 424, 889 (1940); Halverstadt and Kumler, *J. Am. Chem. Soc.*, **64**, 2918 (1942). For general details, see Glasstone, "Recent Advances in Physical Chemistry."
8. Muller, *Phys. Z.*, **33**, 732 (1932); **34**, 689; (1933); Frank, *Proc. Roy. Soc.*, **152A**, 171 (1935); Higasi, *Bull. Inst. Phys. Chem. Res. Tokyo*, **13**, 1167 (1934); *Sci. Pap. Inst. Phys. Chem. Res. Tokyo*, **28**, 284 (1936); **31**, 311, 317 (1937); for reviews, see Glasstone, *Ann. Rep. Chem. Soc.*, **33**, 117 (1936); Smyth, *J. Phys. Chem.*, **43**, 131 (1939).
9. For full list of data, see *Trans. Farad. Soc.*, **30**, appendix (1934).
10. Eucken and Meyer, *Phys. Z.*, **30**, 397 (1929); Smallwood and Herzfeld, *J. Am. Chem. Soc.*, **52**, 1919 (1930); Williams, *ibid.*, **50**, 2350 (1928); Sidgwick and Bowen, *Ann. Rep. Chem. Soc.*, **28**, 367 (1931); Smyth, *op. cit.*; *Chem. Rev.*, **6**, 549 (1929); *J. Am. Chem. Soc.*, **63**, 57 (1941); Remick, *J. Chem. Phys.*, **9**, 653 (1941).
11. For review of applications of dipole moments, see Smyth, *op. cit.*; Le Fèvre, *op. cit.*; Glasstone, *ref. (7)*; Bergmann and Weizmann, *Chem. Rev.*, **29**, 553 (1941).
12. Sutton, *Proc. Roy. Soc.*, **133A**, 668 (1931); *Trans. Farad. Soc.*, **30**, 789 (1934); Kumler and Porter, *J. Am. Chem. Soc.*, **56**, 2549 (1934); Birtles and Hampson, *J. Chem. Soc.*, **10** (1937); Ingles and Hampson, *ibid.*, 981 (1939); Smyth, *et al.*, *J. Am. Chem. Soc.*, **63**, 57 (1941); **64**, 2212 (1942).
13. Gans, *Ann. Phys.*, **64**, 481, **65**, 111 (1921); Stuart, *Z. Phys.*, **55**, 358; **59**, 13 (1929); **63**, 533 (1930); *Ann. Phys.*, **18**, 121 (1933); *Z. phys. Chem.*, **B17**, 429 (1932); *Z. Elek.*, **40**, 478 (1934); *op. cit.*; Hilke, *Z. Phys.*, **103**, 350, 368, 387 (1936); for review, see Beams, *Rev. Mod. Phys.*, **4**, 133, (1932).
14. Thompson, "Chemical Spectroscopy"; Brode, *op. cit.*; Sutherland, *op. cit.*; Barnes and Bonner, *J. Chem. Ed.*, **14**, 565 (1937).
15. For reviews of molecular spectra, see Jevons, "Report on Band Spectra of Diatomic Molecules"; Herzberg, *op. cit.*; Glasstone, "Theoretical Chemistry," Chap. IV and V; *ref. (7)*; Barnes and Bonner, *J. Chem. Ed.*, **14**, 565 (1937); **15**, 25 (1938); Noyes, *ibid.*, **20**, 223 (1943); Randall, *J. App. Phys.*, **10**, 769 (1939); Thompson, *J. Chem. Soc.*, 183 (1944); see also, *Ann. Rep. Chem. Soc.*, **31**, 21 (1934); **32**, 53 (1935); **33**, 53 (1936); **35**, 37 (1938).
16. Franck, *Trans. Farad. Soc.*, **21**, 536 (1925); *Z. phys. Chem.*, **120**, 144 (1926); Birge and Sponer, *Phys. Rev.*, **28**, 259 (1926); Birge, *Trans. Farad. Soc.*, **25**, 707 (1929).
17. For summaries see Sutherland, *op. cit.*; Hibben, *Chem. Rev.*, **15**, 345 (1933); **18**, 1 (1936); "The Raman Effect and its Chemical Applications"; Murphy, *J. Opt. Soc. Am.*, **30**, 396 (1940); Glockler, *Rev. Mod. Phys.*, **15**, 111 (1943).
18. For reviews, see *Ann. Rep. Chem. Soc.*, **28**, 367 (1931); **32**, 53 (1935); **33**, 53 (1936); **35**, 37 (1938); Glasstone, *ref. (15)*.
19. Badger, *J. Chem. Phys.*, **2**, 128 (1934); **3**, 710 (1935); Clark, *Phil. Mag.*, **19**, 476 (1935); **22**, 1137 (1938); *Trans. Farad. Soc.*, **37**, 293, 299 (1941); Allen and Longair, *Phil. Mag.*, **19**, 1032 (1935); Thompson and Linnett, *J. Chem. Soc.*, 1396 (1937); Fox and Martin, *ibid.*, 2106 (1938); Sutherland, *Ann. Rep. Chem. Soc.*, **35**, 46 (1938); *J. Chem. Phys.*, **8**, 161 (1940); Burawoy, *Trans. Farad. Soc.*, **39**, 79 (1943); Wu and Yang, *J. Phys. Chem.*, **48**, 295 (1944).
20. Brode, *op. cit.*
21. Bury, *J. Am. Chem. Soc.*, **57**, 2115 (1935); Sklar, *J. Chem. Phys.*, **5**, 669 (1937); **10**, 521 (1942); Schwarzenbach, *et al.*, *Helv. Chim. Acta*, **31**, 1636 (1938); Lewis and Calvin, *Chem. Rev.*, **25**, 273 (1939); Brode, *J. App. Phys.*, **10**, 751 (1939); *op. cit.*; Förster, *Z. Elek.*, **45**, 548 (1939); Pauling, *Proc. Nat. Acad. Sci.*, **25**, 577 (1939); Rodebush, *et al.*, *J. Am. Chem. Soc.*, **62**, 2906 (1940); **63**, 3018 (1941); Sklar, *Rev. Mod. Phys.*, **14**, 232 (1942); Booker, *ibid.*, p. 275; Herzfeld and Sklar, *ibid.*, p. 294; Sheppard, *ibid.*, p. 303; *J. Am. Chem. Soc.*, **64**, 2922, 2937 (1942); Lewis, *et al.*, *ibid.*, **65**, 2102, 2107 (1943); Jones, *ibid.*, pp. 1815, 1818; *Chem. Rev.*, **32**, 1 (1943); Bowen, *Ann. Rep. Chem. Soc.*, **40**, 12 (1943).

22. Hartley and Dobbie, *J. Chem. Soc.*, **75**, 640 (1899); Baly and Desch, *ibid.*, **85**, 1029 (1905); Hantzsch, *Ber.*, **43**, 3049 (1910); **44**, 1771 (1911); Brode, *J. Am. Chem. Soc.*, **56**, 581 (1924); Calvin and Wood, *ibid.*, **62**, 3152 (1940); Zechmeister, *et al.*, *ibid.*, **65**, 1522, 1940 (1943).
 23. von Halban and Eisenbrand, *Z. phys. Chem.*, **132**, 401, 433 (1928); Wesp and Brode, *J. Am. Chem. Soc.*, **56**, 1037 (1934); Macwater and Barratt, *J. Chem. Soc.*, 517 (1934); Bent and French, *J. Am. Chem. Soc.*, **63**, 568 (1941); Edwards and Birnbaum, *ibid.*, p. 1471; Rabinowitch and Stockmayer, *ibid.*, **64**, 335 (1942); Gould and Vosburgh, *ibid.*, p. 1630.
 24. Fajans, *Ber.*, **53**, 643 (1920); **55**, 2826 (1922); Eucken, *Annalen*, **440**, 111 (1924); Grimm, *Z. phys. Chem.*, **102**, 134 (1922); *Z. Elek.*, **31**, 474 (1925); Rossini, *Bur. Stand. J. Res.*, **13**, 21 (1934); Taylor and Smith, *J. Chem. Phys.*, **7**, 390 (1939); **8**, 543 (1940); Burton, *et al.*, *ibid.*, pp. 297, 743; Stevenson, *ibid.*, **10**, 291 (1942); Kistiakowsky, *et al.*, *ibid.*, p. 305; **11**, 6 (1943); Herzberg, *ibid.*, **10**, 306 (1942); Van Artsdalen, *ibid.*, p. 653; Pauling, *op. cit.*, Chap. II; Rice, "Electronic Structure and Chemical Binding," Chap. XII.
 25. Pauling and Sherman, *J. Chem. Phys.*, **1**, 606, 679 (1933); Kistiakowsky, *et al.*, *J. Am. Chem. Soc.*, **58**, 137, 146 (1936); Pauling, *op. cit.*, Chap. II and VI.
 26. Mark and Wierl, *Z. Phys.*, **60**, 741 (1930); *Z. Elek.*, **36**, 375 (1930); Wierl, *Phys. Z.*, **31**, 366, 1028 (1930); Pauling and Brockway, *J. Chem. Phys.*, **2**, 867 (1934); *J. Am. Chem. Soc.*, **57**, 2684 (1935); **59**, 1223 (1937); Bauer, *J. Chem. Phys.*, **4**, 406 (1936); Pirenne, *ibid.*, **7**, 144 (1939); Walter and Beach, *ibid.*, **8**, 601 (1940). For reviews of methods and results, see Brockway, *Rev. Mod. Phys.*, **8**, 231 (1936); Brockway and Taylor, *Ann. Rep. Chem. Soc.*, **34**, 196 (1937); Glasstone, *ibid.*, **33**, 65 (1936); Maxwell, *J. Opt. Sci. Am.*, **30**, 374 (1940); Spurr and Pauling, *J. Chem. Ed.*, **18**, 458 (1941).
 27. Lowry, *Trans. Farad. Soc.*, **10**, 96 (1914); *op. cit.*
 28. Tschugaeff, *Z. phys. Chem.*, **85**, 507 (1913); Lowry and Gore, *J. Chem. Soc.*, 319 (1931); Lowry and Hudson, *Phil. Trans. Roy. Soc.*, **232A**, 117 (1933); Kuhn, *Z. phys. Chem.*, **B8**, 281 (1930); *Ber.*, **63**, 190 (1930); see discussion on Optical Rotatory Power, *Trans. Farad. Soc.*, **26**, 266, *et seq.* (1930).
 29. Hudson, *J. Am. Chem. Soc.*, **31**, 66 (1909); Read, *Trans. Farad. Soc.*, **26**, 441 (1930); Lowry, *et al.*, *J. Chem. Soc.*, 1179 (1933); Kauzmann, *J. Am. Chem. Soc.*, **64**, 162 (1942).
 30. Beckmann and Cohen, *J. Chem. Phys.*, **4**, 784 (1936); **6**, 163 (1938); **8**, 827, 831 (1940); Pryde and Rule, *J. Chem. Soc.*, 345 (1940); Eyring, *et al.*, *Chem. Rev.*, **26**, 339 (1940); see also, Rule, *Trans. Farad. Soc.*, **26**, 321 (1930); Betti, *ibid.*, p. 337.
 31. Cotton, *Ann. Chim. Phys.*, **8**, 347 (1896); Kuhn, *loc. cit.*, ref. (28); *Trans. Farad. Soc.*, **26**, 293 (1930); Lowry, *op. cit.*
 32. For reviews, see Allsopp, *Ann. Rep. Chem. Soc.*, **32**, 120 (1935); Lowry, *op. cit.*; Condon, *Rev. Mod. Phys.*, **9**, 432 (1937); Kauzmann, Walter and Eyring, *Chem. Rev.*, **26**, 339 (1940).
 33. Pauling, *Proc. Roy. Soc.*, **114A**, 181 (1927); Angus, *ibid.*, **136A**, 569 (1932); Brindley and Hoare, *ibid.*, **152A**, 342 (1935); *Trans. Farad. Soc.*, **33**, 268 (1937); Trew, *ibid.*, **37**, 476 (1911); Klemm, *Z. anorg. Chem.*, **246**, 347 (1911).
 34. Pascal, *Compt. rend.*, **149**, 342, 508 (1909); *Bull. Soc. Chim.*, **5**, 758, 859 (1909); **7**, 17, 45 (1910); for review, see Angus, *Ann. Rep. Chem. Soc.*, **38**, 27 (1911).
 35. Stoner, *Phil. Mag.*, **8**, 250 (1929); Pauling, *J. Am. Chem. Soc.*, **53**, 1367 (1931).
 36. Taylor, *ibid.*, **48**, 854 (1926); Sugden, *Trans. Farad. Soc.*, **30**, 18 (1934); *J. Chem. Soc.*, 440 (1936); Müller, *Annalen*, **520**, 230, 235; **521**, 81, 89 (1935); *Ber.*, **72**, 2063 (1939); **74**, 807, 1051, 1075 (1941).
- For reviews of magnetochemistry, see Klemm, *Z. Elek.*, **45**, 583 (1939); Müller, *ibid.*, p. 593; Selwood, *op. cit.*; *J. Chem. Ed.*, **19**, 181 (1912); Sugden, *J. Chem. Soc.*, 328 (1943).
37. Welo and Baudisch, *Nature*, **116**, 606 (1925); Bosc, *Phys. Z.*, **43**, 864 (1927); Pauling, *ref. (35)*; *op. cit.* Chap. III; Van Vleck, *J. Chem. Phys.*, **3**, 807 (1935).

CHAPTER IX

DILUTE SOLUTIONS

COLLIGATIVE PROPERTIES

In the present chapter it is proposed to consider the so-called colligative properties (p. 524) of liquid solutions, namely, those depending mainly on the number, and not the nature, of the molecules present. The particular properties to be discussed are sometimes known as osmotic properties, for they include, in addition to the osmotic pressure, the lowering of vapor pressure, rise of boiling point and depression of the freezing point of solutions, which can be related theoretically to the osmotic pressure. In general, a **solution** may be described as a homogeneous mixture, constituting one phase only, of two or more components, and although there is no fundamental difference between the rôles of these components (cf. p. 710), it is convenient, for certain purposes, to call the one present in excess the **solvent** and the other, i.e., the dissolved substance, the **solute**. For the purpose of the discussion of the colligative properties of solutions, the solute is assumed to be nonvolatile, whereas the solvent is volatile.

Lowering of Vapor Pressure by Dissolved Substances.—Whenever a substance is dissolved in a liquid the vapor pressure of the latter is lowered; this fact has been known qualitatively for many years, and was even studied quantitatively at the beginning of the 19th century, e.g., C. L. Berthollet, 1803; M. Faraday, 1822. The first important contribution to the subject was made by L. von Babo (1847); he measured the vapor pressures of aqueous salt solutions at a number of temperatures, and concluded that for any given solution the decrease of vapor pressure, from that of the pure solvent, is a constant fraction of the latter at all temperatures. If p^0 is the vapor pressure of the pure solvent and p that of the solution, then according to von Babo $(p^0 - p)/p^0$, now known as the **relative lowering** of vapor pressure, is a constant for any given solution, independent of the temperature at which it is measured. This result has been verified experimentally by other workers, but it is, theoretically, strictly true under certain conditions only. The Clapeyron-Cläusius equation, in the form applicable to the system liquid-vapor, the latter behaving as an ideal gas (p. 452), may be written

$$\frac{d \ln p^0}{dT} = \frac{L_o^0}{RT^2} \quad (1)$$

for the pure solvent; for the solution it becomes

$$\frac{d \ln p}{dT} = \frac{L_e}{RT^2}, \quad (2)$$

where L_0^0 and L_* are the amounts of heat required to vaporize a mole of solvent from pure solvent and from the solution, respectively. Subtracting (1) from (2), it follows that

$$\frac{d \ln (p/p^0)}{dT} = \frac{L_* - L_0^0}{RT^2}. \quad (3)$$

By supposing 1 mole of solvent to be vaporized from the pure solvent, the heat absorbed being L_0^0 , and then condensed into the solution, the volume of which is so large that its concentration is not thereby appreciably altered, so that the heat evolved is L_* , it is seen that the net heat change $L_* - L_0^0$ is equal to the differential heat of dilution of the solution (p. 242). Since the relative lowering of the vapor pressure is $1 - p/p^0$, it follows from (3) that it can only be independent of temperature if the heat of dilution of the solution is zero. This can only be true for an ideal solution (see p. 687), but for dilute solutions the heat of dilution is sufficiently small for the relative lowering of vapor pressure to be almost independent of temperature. It can be deduced from equation (3) that if heat is absorbed on dilution, as is generally true for solutions of appreciable concentration, then the relative lowering of vapor pressure must increase as the temperature is raised.

Although von Babo was unable to discover any simple relationship between the lowering of the vapor pressure and the amount of dissolved substance, A. Wüllner (1856-60), also working with solutions of salts in water, concluded that the vapor pressure lowering was proportional to the concentration of the solution. The studies of F. M. Raoult (1887-88) in this connection marked a significant advance and led to a generalization of fundamental importance to the study of solutions. This investigator confirmed the work of von Babo and of Wüllner, and from measurements with a variety of solutes in a number of organic solvents, he finally proposed the relationship

$$\frac{p^0 - p}{p^0} = \frac{n_2}{n_1 + n_2} = x_2, \quad (4)$$

where n_1 and n_2 are the number of moles of solvent and solute respectively, and x_2 is the mole fraction of solute.* The results in Table 102, taken

TABLE 102. TEST OF RAOULT'S EQUATION IN ETHYL SOLUTION

Solute	x_2	$\frac{p^0 - p}{p^0}$	$\frac{p^0 - p}{p^0 x_2}$
Nitrobenzene	0.060	0.0554	0.92
Methyl salicylate	0.092	0.086	0.94
Ethyl benzoate	0.096	0.091	0.95
Benzaldehyde	0.130	0.132	1.02
Aniline	0.077	0.081	1.06

* For solutions of the type discussed in this chapter the convention used earlier (p. 532), whereby the solvent is designated by the subscript 1 and the solute by 2, is employed; when it is unnecessary to distinguish between the two components of the solution the subscripts A and B are used (see p. 704).

from the work of Raoult, show that (4) is obeyed, approximately, for a number of solutes in ether solution. For the Raoult equation to hold exactly the results in the last column should have been unity, but in view of experimental difficulties, and the fact that the solutes were themselves volatile, the agreement may be regarded as satisfactory. More recent accurate measurements have confirmed the reliability of (4), often even in relatively concentrated solutions; some of the precise data obtained for aqueous solutions of mannitol at 20° c. (J. C. W. Frazer, *et al.*, 1920) are recorded in Table 103. The measured values of $p^0 - p$ are

TABLE 103. VAPOR PRESSURE LOWERING OF AQUEOUS SOLUTIONS OF MANNITOL AT 20° C.

G. Mannitol to 1000 g. Water	Moles Mannitol to 1000 g. Water	$p^0 - p$	
		Obs.	Calc.
17.930	0.0984	0.0307	0.0311
36.004	0.1977	0.0614	0.0622
71.917	0.3945	0.1227	0.1239
108.243	0.5944	0.1860	0.1858
126.283	0.6934	0.2162	0.2164
162.332	0.8913	0.2791	0.2775
180.451	0.9908	0.3096	0.3076

compared with those calculated by (4); the vapor pressure p^0 of pure water at 20° c. is taken as 17.54 mm. of mercury.¹

By subtracting unity from both sides of (4), it is seen that

$$\frac{p}{p^0} = 1 - x_2 = x_1;$$

$$\therefore p = p^0 x_1, \quad (5)$$

where x_1 is the mole fraction of the solvent, since $x_1 + x_2$ equals unity. It follows, therefore, that the vapor pressure of the solvent in a solution obeying Raoult's equation is directly proportional to the mole fraction of the solvent; (4) and (5) are thus to be regarded as alternative methods of expressing what has become known as **Raoult's law**.*

Determination of Molecular Weights.—By means of the Raoult equation measurements of lowering of vapor pressure may be used to determine molecular weights. If w_2 g. of solute of molecular weight M_2 are dissolved in w_1 g. of solvent of molecular weight M_1 , then the respective numbers of moles of each are $n_2 = w_2/M_2$ and $n_1 = w_1/M_1$, so that Raoult's law [equation (4)] gives

$$\frac{p^0 - p}{p^0} = \frac{w_2/M_2}{w_1/M_1 + w_2/M_2}. \quad (6)$$

For a dilute solution n_2 may be neglected in comparison with n_1 , so that (6) becomes

$$\approx \frac{w_2}{M_2} \cdot \frac{M_1}{w_1}. \quad (7)$$

* In the idealized form of Raoult's law the (observed) vapor pressures are replaced by the corresponding fugacities (cf. p. 301).

By making up a solution of known weights of solvent and solute it is possible, therefore, to determine the molecular weight of the latter, if the vapor pressure of solvent and solution can be measured. In view of the small value of the relative lowering and the difficulties involved in its determination the method is not commonly employed, although by the use of certain devices, described below, it is often possible to obtain approximate results with relative ease.

Measurement of Vapor Pressure Lowering: (i) Differential Methods.—

The earliest observations were made by the static method involving a barometer tube (p. 446), but this is unsatisfactory partly because traces of air and other impurities present in the liquid introduce serious errors, and partly because vapor pressure lowerings are very small at ordinary temperatures. In a molar aqueous solution, for example, the mole fraction of solute is about 0.018, and so at 25° c., when p^c is 23.76 mm. of mercury, the lowering of vapor pressure would be given by

$$p^0 - p = p^0 x_2 = 23.76 \times 0.018 = 0.4277 \text{ mm.}$$

To avoid the necessity of estimating these small pressure differences, Raoult generally made measurements at such temperatures as to give vapor pressures of about 200 to 400 mm., so that the lowering was increased proportionately. For accurate work, however, it is desirable to make observations at specific temperatures, and consequently various methods have been proposed for the determination of small pressure differences. One such procedure involves the use of an accurate differential manometer. In the method of C. Dieterici (1893) the saturated vapors of solvent and solution were exposed to the two sides of a thin glass disc of 7 cm. diameter and 0.08 mm. thickness. The difference of pressure caused the disc to move and the movement was magnified, and thus

measured, by a beam of light reflected from a mirror. After calibration, with the aid of another manometer of lower sensitivity, it was claimed that the apparatus was capable of measuring pressure differences as small as 0.01 to 0.1 mm. of mercury with an accuracy of 0.001 mm. Other forms of differential manometer were used for vapor pressure determinations by A. Smits (1902) and O. F. Tower (1908), but special interest attaches to the apparatus of J. C. W. Frazer and B. F. Lovelace (1914), generally regarded as giving the most accurate results in this field, which utilizes the principle of the Rayleigh manometer. Into the bulbs A and B (Fig. 133), about 4 cm. in diameter and connected by a movable mercury reservoir attached to C, are sealed the glass points PP which are associated with the mirror M firmly mounted in a plane perpendicular to that of the rest of the apparatus. By connecting the side tubes D and E the pressure is made the same in the two bulbs, and by means of a fine screw the mercury reservoir is adjusted so that the

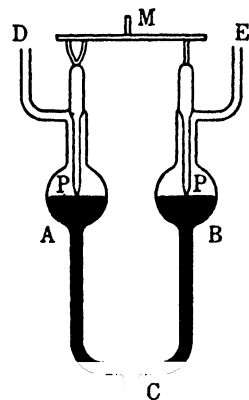


FIG. 133. Lowering of vapor pressure (Frazer and Lovelace).

points PP correspond exactly with their images in the two mercury surfaces; this gives the zero reading of the manometer. One of the bulbs is now con-

nected, by its side tube, to the vapor of the given solution and the other to the pure solvent; there is consequently a difference in the two mercury levels, but by tilting the apparatus gradually it is possible, with the aid of the reservoir, to make the points *PP* coincide again with their images. The extent of the tilt, measured by a beam of light reflected from *M*, is a measure of the pressure difference in the two bulbs. By adopting careful temperature control and taking special precautions to remove all traces of air from the solvent and solutions, very accurate results, with an error of less than 0.0006 mm. of mercury, can be obtained.

Differences of vapor pressure can be determined, less accurately, by the tensimeter of P. C. F. Frowein (1887), or by a modified apparatus depending on the same principle. The differential manometer consists of a U-tube containing a liquid of low volatility and relatively low density; at one time β -bromonaphthalene was commonly used, but in recent years *n*-butyl phthalate or Apiezon oil have been generally employed. The vessel containing the solution is attached to one limb of the manometer, and one of pure solvent to the other: the two vessels are first connected together and the apparatus evacuated as completely as possible, precautions being taken to remove dissolved air. The connection between the vessels is closed off, and the difference in vapor pressure measured by the difference in level of the two limbs of the manometer when equilibrium is reached. R. E. Gibson and L. H. Adams (1933) claim an accuracy in reading pressure differences by this method equivalent to ± 0.01 mm. of mercury.

A device for making the solution act as its own manometric liquid is contained in the apparatus described by A. W. C. Menzies (1910) for the purpose of determining molecular weights. The vapor pressure difference is measured at the boiling point so that $p^0 - p$ is much greater than at room temperature. Solvent, free from air, is placed in the bulb *A* (Fig. 134) and in the inner graduated tube *B*; the bulb *A* is heated and by closing the clamp *D* and lifting the ground-glass stopper *E*, vapor is forced through the tube *C* into the liquid in *B* so that it is raised to the boiling point. The clamp is now slowly opened and the stopper lowered into place. Since *A* and *B* both contain boiling solvent, the levels in *B* and *C* should be the same, but it is actually slightly higher in the latter owing to capillarity, and the difference is read off from the graduations on the tube *C*. The heating is now stopped, a weighed quantity of solute added to the liquid in *B* and the procedure of raising the temperature to that of the boiling solvent is repeated. The pressure of the vapor in *A*, containing pure solvent, is now greater than that in *B*, containing solution, and so the level of the liquid in *C* will be depressed below that of the surrounding solution. This difference of level, after making the correction for capillarity, gives $p^0 - p$ in mm. of solution at the boiling point, and this may be readily converted into mm. of mercury. The value of p^0 is the vapor pressure of the solvent at its boiling point, and hence is equal to the barometric pressure. In order to calculate the molecular weight of the solute it is now only necessary to

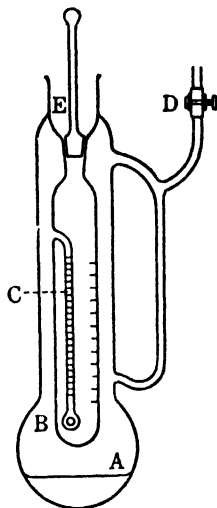


FIG. 134. Lowering of vapor pressure (Menzies)

know the weight of the solvent, and this may be obtained with sufficient accuracy by reading off the volume of solution in B .²

(ii) **Dynamic Method.**—The dynamic method of measuring the vapor pressure of a pure liquid (p. 447) has been adapted for the study of solutions. The temperature of the liquid, viz., pure solvent or solution, is kept constant and the pressure adjusted until the liquid boils; the particular pressure at which this occurs is equal to the required vapor pressure at the given temperature. The method was used by A. A. Noyes and C. G. Abbott (1897), H. M. Goodwin and G. K. Burgess (1899), and B. C. Burt (1904), but it does not appear to have been employed to any great extent in recent years. The chief sources of error are superheating (cf. p. 638) and change of concentration of the solution due to evaporation. It may be noted that, in a sense, the method of Menzies, described above, represents a differential form of the dynamic method, the solvent boiling at atmospheric pressure and the solution at a lower pressure at the same temperature.

(iii) **Transpiration Method.**—The transpiration, or gas saturation, method (p. 448), which has been widely used in the accurate form described below, was first applied to the study of solutions by J. Walker (1888) at the suggestion of W. Ostwald. Dry air was passed in succession through a set of Liebig bulbs containing the solution, then through similar bulbs containing the pure solvent (water) and then into absorption tubes filled with calcium chloride. In passing through the first bulbs the initially dry air became saturated with water vapor to a pressure p , and so these bulbs decreased in weight by an amount proportional to p . The air then entered the solvent bulbs where the pressure of vapor was raised to p^0 , and so the decrease in weight in these was proportional to $p^0 - p$. In the absorption tubes all the water was removed, the water vapor pressure was thus reduced from p^0 to zero, and so the gain in weight was proportional to p^0 . The ratio of the loss in weight of the second set of bulbs to the gain in weight of the absorption tubes is thus equal to $(p^0 - p)/p^0$, the relative lowering of vapor pressure. Although this method can be used for the approximate determination of molecular weights, it does not give accurate results in the simple form described. The chief sources of error are (i) the volume of air drawn through the two sets of bulbs would not be the same, as shown below, even if the total pressure remained constant, although there are actually variations of pressure throughout the apparatus as used; and (ii) the process of saturation by bubbling introduces errors because (a) spray is mechanically carried over from the solution to the solvent, and (b) the decrease in pressure in a bubble as it rises through the liquid causes expansion and there is a possibility of the initially saturated bubble becoming unsaturated. The Earl of Berkeley and E. G. J. Hartley (1906) designed saturators to avoid these difficulties, the air being led slowly over the surface of the liquid kept in a state of oscillation, and similar apparatus has been used later by others.

A considerable improvement in technique was made by E. W. Washburn and E. W. Heuse (1915) who claimed an accuracy of 0.5 per cent for their determinations of relative lowering of vapor pressure. The air is first slightly supersaturated with respect to water at the experimental temperature, e.g., 25° c., by passing through a saturator at a higher temperature; it is then passed through a saturator I containing pure water at 25° c., and on to absorber I where all the water vapor is removed. The dry air next goes to a pre-saturator, where it becomes nearly saturated, then through saturator II containing the

experimental solution at 25° c., and to another absorber II. The use of the pre-saturator ensures complete saturation in II, and also prevents an appreciable change in concentration of the solution. Finally, the air passes through a third saturator III containing pure water at 25° c. and to an absorber III. In saturators I and III equilibrium with the solvent is approached from over- and under-saturation, respectively, so that the results from I and III provide a check on each other. The arrangement is as shown below:

→Supersaturator→Saturator I→Absorber I→Pre-saturator→
(Pure solvent)

Saturator II→Absorber II→Saturator III→Absorber III→
(Solution) (Pure solvent)

The calculations are carried out as follows. If Dalton's law of partial pressures applies to a mixture of n_A moles of dry air and n_B moles of water vapor, then the respective partial pressures p_A and p_B are related by (cf. p. 303)

$$n_A/n_B = p_A/p_B. \quad (8)$$

If the gas laws apply to the air, then $p_A v_A = n_A RT$, where v_A is the volume of n_A moles of air at the partial pressure p_A , and combination with (8) gives

$$n_B = p_B v_A / RT. \quad (9)$$

From this it is evident that the number of moles, and hence the mass m , of water required to saturate a volume v_A of air to a definite pressure p is given by

$$m = k p v_A, \quad (10)$$

where k is a constant at a given temperature. In the method described above, the same weight of air is passed through the various saturators and consequently the volume of air, apart from water vapor, emerging from any saturator will be inversely proportional to its partial pressure in that saturator; this is equal to the total pressure (B) in the saturator minus the vapor pressure (p) of the liquid in the saturator. For two saturators I and II it follows, therefore, that

$$\frac{v_I}{v_{II}} = \frac{B_{II} - p_{II}}{B_I - p_I}. \quad (11)$$

Since the mass of water required to saturate a particular volume of air is given by (10), it follows that

$$\frac{m_I}{m_{II}} = \frac{p_I v_I}{p_{II} v_{II}}, \quad (12)$$

where m_I and m_{II} are the masses of water taken up by the absorbers I and II, which immediately follow the saturators I and II. From (11) and (12) it can be shown that

$$\frac{p_I - p_{II}}{p_{II}} = \frac{m_I(B_I - p_I) - m_{II}(B_{II} - p_I)}{m_{II}B_{II}} = A. \quad (13)$$

Since saturators I and II contain solvent and solution respectively, p_I and p_{II} are equivalent to p^0 and p , and so the relative lowering of the vapor pressure is given by

$$\frac{p^0 - p}{p^0} = \frac{A}{1 + A}, \quad (14)$$

where A is defined by (13). To determine A it is necessary to know, in addition

to the gain in weight of the absorbers, the vapor pressure of pure water p_1 , generally obtained from tables, and the total pressure in the two saturators. The difference between the pressure in the saturator and that of the atmosphere, as given by the barometer, can be measured by a simple U-tube form of manometer, one limb being connected to the outlet end of the saturator and the other open to the atmosphere. The transpiration method of J. N. Pearce and R. D. Snow (1927), described on p. 448, has been extensively used for studies with aqueous solutions. The only additional feature which need be mentioned here is that pre-saturators are used, as described above, in order to facilitate saturation of the air by the solution. It should be emphasized that the accuracy of air-saturation methods depends ultimately on the applicability of Dalton's law of partial pressures to the mixture of air and water vapor.³

(iv) **Dew-Point Method.**—When a vapor at a given pressure is cooled, liquid just commences to form at the temperature for which that pressure is the saturation value; this principle is utilized in the **dew-point method** for determining vapor pressures, proposed by H. Lescoeur (1889) and improved by A. C. Cumming (1906). The aqueous solution is placed in *A* (Fig. 135), and through the rubber stopper passes a polished silver tube *B* fitted with a thermometer and inlet and outlet tubes. The apparatus is placed in a thermostat and evacuated by connecting *C* to a pump; the stopcock is closed and the system allowed to reach equilibrium. A quantity of ether is put into *B* and a slow stream of air passed through it, so as to cause the temperature to fall gradually; at a certain point dew forms on *B*, detected by the sudden dimming of the polished silver, and the temperature of the tube is then read on the thermometer. The saturation vapor pressure of water at this temperature is obtained from tables, and this represents the pressure of water vapor existing in *A*; it is consequently also the vapor pressure of the solution in *A* at the temperature of the thermostat. In the modified procedure of J. W. McBain and C. S. Salmon (1920), used in the study of the vapor pressures of soap solutions (cf. p. 1266) at 90° c., a better control of the temperature of the silver tube was obtained by maintaining a flow of water through it from a regulated reservoir. The temperature of the latter was allowed to fall slowly and the formation of dew observed; the temperature of the reservoir was then raised gradually and the point at which the dew disappeared was also recorded. In this way an accuracy of 0.01° was obtained in the measurement of the dew point.⁴

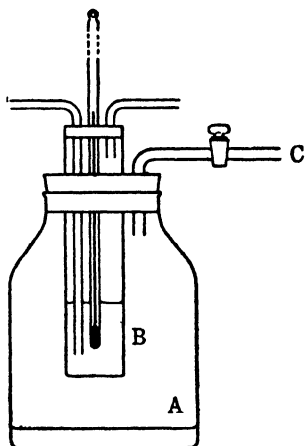


FIG. 135. Vapor pressure by dew-point method

(v) **Isopiestic Methods.**—If two vessels containing different solutes in the same solvent are placed side by side in a closed space, vapor will distil from the solution of higher vapor pressure and condense in the one having the lower pressure until, when equilibrium is attained, both solutions are exerting the same vapor pressure, that is to say, they are **isopiestic** * (Greek: *equal pressure*). The solutions can then be analyzed to determine their concentrations, and if one of the solutes is potassium chloride or mannitol, the vapor pressures for whose

* They are also referred to as "isotonic" solutions (see p. 661).

solutions are known at various concentrations, the vapor pressure of the isopiestic solution can be determined. One of the chief difficulties in the application of this principle (W. R. Bousfield, 1917) is the slowness with which the solutions reach equilibrium, but this has been overcome (D. A. Sinclair, 1933; R. A. Robinson and D. A. Sinclair, 1934). Four silver dishes, gold-plated internally, 3.8 cm. square and 2 cm. deep, are placed in close contact on a thick copper block, in a desiccator which can be rocked in a thermostat. A potassium chloride solution is introduced into two of the dishes and solutions of the substance under investigation are placed in the others. The desiccator is evacuated and the system allowed to come to equilibrium; the solutions are then removed and their concentrations determined. To ensure rapid attainment of equilibrium the potassium chloride solutions are chosen so as to have vapor pressures close to those expected for the other solutions. It has been claimed that this method is capable of giving vapor pressures at least as accurate as those obtained in any other manner.

Several methods, involving the isopiestic principle, have been proposed for the approximate determination of molecular weights: that of G. Barger (1904) is of interest as only very small amounts of material are required. A solution of definite concentration of a substance of known molecular weight is prepared, and a series of solutions of the substance under investigation are made in the same solvent at known concentrations. Alternate drops, separated by small air spaces, of the former and of one of the latter solutions are introduced into a capillary tube 6–8 cm. long and 1.5 mm. bore, and observed under the microscope. In this arrangement, each drop of the solution of the unknown is separated on each side by a small air space from a drop containing a substance of known molecular weight at a definite concentration. If the vapor pressure of the former is greater than that of the latter, then as a result of the tendency for the solutions to become isopiestic the drops of the latter will increase at the expense of the former, and *vice versa*. By a few experiments it is possible to find a solution of the substance of unknown molecular weight, the drops of which remain unchanged; it is then isopiestic with the solution containing the known solute. Since the vapor pressure is the same for both solutions, and the solvent is the same, the relative lowerings of the vapor pressure are equal. If Raoult's law applies to both solutes then their respective mole fractions are the same, and if the solutions are not too concentrated the molar concentrations may be taken as approximately equal. Since the molecular weight of one substance is known, that of the other can be calculated.⁵

ELEVATION OF THE BOILING POINT

Vapor Pressure and Boiling Point.—A direct consequence of the reduction of vapor pressure by a non-volatile solute is that the boiling point of the solution, i.e., the temperature at which its vapor pressure is equal to the atmospheric pressure, must be higher than that of the pure solvent. Studies of the rise of boiling point produced by dissolved substances were made by M. Faraday (1820) and by others in the early part of the 19th century, but it is to F. M. Raoult (1871, 1889) and especially to E. Beckmann (1889–1914) that credit must be given for accurate measurements. It is of interest to record that Beckmann started the study of solutions by determining the lowering of vapor pressure, but abandoned this in favor

of the elevation of boiling point, to which it is obviously related, as being measurable with a greater degree of accuracy. The rise of the boiling point of a solution has subsequently become of special interest as it provides a relatively simple method for the determination of molecular weights of dissolved substances.

In Fig. 136 are two curves showing the variation of the vapor pressure with temperature of pure solvent and a solution; the latter is seen to be

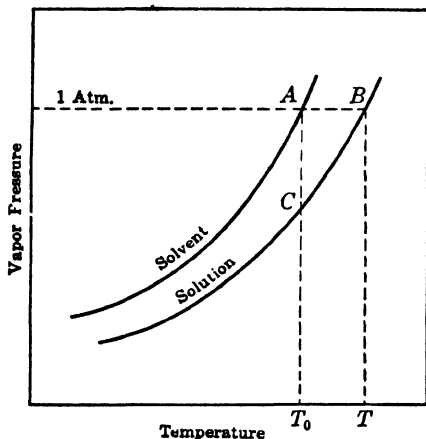


FIG. 136. Elevation of the boiling point

always below the former. The boiling point of the solvent is T_0 and that of the solution is T , so that the elevation of the boiling point $T - T_0 = \Delta T_e$ is represented by the distance AB in the diagram. The lowering of the vapor pressure $p^0 - p$ at the temperature T_0 is equivalent to AC , and since p^0 , the vapor pressure of the pure solvent, is then a constant, viz., 1 atm., it is evident that AC is proportional to $(p^0 - p)/p^0$ at this temperature. If the vapor pressure curves for a series of solutions may be regarded as practically parallel in the region of the boiling point, as is probably the case for

dilute solutions, then AB/AC will be constant for these solutions. In other words, the rise of boiling point will be proportional to the relative lowering of vapor pressure. Further, since the latter is equal to the mole fraction x_2 of the solute [equation (4)], it follows that

$$\Delta T_e = k_e x_2, \quad (15)$$

where k_e is a constant. For a dilute solution the approximate form, as in (7), may be used for x_2 , so that

$$\Delta T_e = k_e \frac{w_2}{M_2} \cdot \frac{M_1}{w_1}. \quad (16)$$

Since w_2/M_2 is the number of moles of solute dissolved in w_1 g. of solvent, it follows, k_e and M_1 being constants, that the rise of boiling point in a given solvent is approximately proportional to the molar concentration of the solute, and should be independent of its nature. The elevation of the boiling point is thus a colligative property.

Thermodynamics of Elevation of Boiling Point.—The result obtained above may be deduced in a more precise manner by means of thermodynamics; these deductions were first made, along somewhat different

lines, by J. H. van't Hoff (1886) for the elevation of boiling point and other colligative properties.* Making the usual assumptions (cf. p. 452), the Clapeyron-Clausius equation for the solution may be written

$$\frac{d \ln p}{dT} = \frac{L_o}{RT^2},$$

where L_o is latent heat of evaporation of 1 mole of solvent from the solution. Assuming this heat quantity to remain constant, integration between the temperatures T_0 and T , the boiling points of solvent and solution, respectively, gives *

$$\ln \frac{p}{p^0} = -\frac{L_o}{R} \left(\frac{1}{T_0} - \frac{1}{T} \right) \quad (17)$$

$$= -\frac{L_o}{R} \left(\frac{T - T_0}{TT_0} \right) = -\frac{L_o}{R} \cdot \frac{\Delta T_o}{TT_0}. \quad (18)$$

For dilute solutions, T and T_0 are not very different and so TT_0 may be replaced by T_0^2 and L_o may be taken as the molar heat of evaporation of the pure solvent. Further, according to Raoult's law [equation (5)], p/p^0 is equal to x_1 , the mole fraction of solvent, and hence to $1 - x_2$, where x_2 is the mole fraction of solute; hence (18) may be written

$$\ln (1 - x_2) = -\frac{L_o}{R} \cdot \frac{\Delta T_o}{T_0^2}. \quad (19)$$

If the solution is assumed to be dilute then x_2 is small and $\ln (1 - x_2) \approx -x_2$, so that (19) becomes

$$x_2 = \frac{L_o}{R} \cdot \frac{\Delta T_o}{T_0^2} \quad (20)$$

or

$$\Delta T_o = \frac{RT_0^2}{L_o} x_2. \quad (21)$$

A further consequence of the assumption that the solution is dilute is that x_2 may be put equal to $w_2 M_1 / M_2 w_1$, and hence

$$\Delta T_o = \frac{RT_0^2}{L_o} \cdot \frac{w_2 M_1}{M_2 w_1} \quad (22)$$

$$= \frac{RT_0^2}{l_o} \cdot \frac{w_2}{M_2 w_1}, \quad (23)$$

where $l_o = L_o/M_1$ is the latent heat of vaporization *per gram* of solvent.

* It should be noted that the vapor pressure of the solution is p at T , and p^0 at T (see Fig. 136), where p^0 and p are also the vapor pressures of solvent and solution, respectively, at T_0 .

Equation (23) may also be written in the form

$$\Delta T_0 = \frac{RT_0^2}{1000l_0} \cdot \frac{1000w_2}{M_2w_1} \quad (24)$$

$$= K_0 m, \quad (25)$$

where K_0 , which is equal to $RT_0^2/1000l_0$, is a constant for each solvent, since it depends on its boiling point and latent heat, in addition to universal constants. The quantity m , equal to $1000 w_2/M_2w_1$, is the **molality** of the solution; it is the number of moles of solute dissolved in 1000 g. of solvent. It is evident from (25) that for a dilute solution obeying Raoult's law, the rise of boiling point ΔT_0 is proportional to the molality of the solution.* The proportionality constant K_0 is called the **molal elevation constant**, or the **ebullioscopic constant**, and from (25) it can be seen to be equivalent to the boiling point elevation for a solution of unit molality. Actually such a solution would not be sufficiently dilute for (25) to be applicable, and so K_0 may be regarded as the rise of boiling point for a unit molal solution, if the value were proportional to that for a dilute solution.

There are three aspects of the theoretical treatment which require experimental verification: (i) the elevation of boiling point should be proportional to the molality of the solution, (ii) the proportionality constant K_0 should be independent of the nature of the solute, and (iii) the value of K_0 obtained from experiments on the rise of boiling point should be equal to that calculated from the latent heat. It can be stated that in general the theoretical deductions have been substantiated in practice, but it must be remembered first, that the equations developed can be expected to apply only to dilute solutions, and second, they can hold only for systems to which Raoult's law is at least approximately applicable, since this law has been assumed in the deduction. Observations to verify the theoretical treatment are best made with nonvolatile solutes which are not associated in solution, and which do not form compounds with the solvent.

The results of many experiments show that in dilute solutions the boiling point elevation is very nearly proportional to the concentration or molality. Further, the proportionality constant does not depend on the chemical nature of the solute, provided it is not of a type for which obvious deviations are to be expected. For example, the following values of

Naphthalene	2.07	Phenyl benzoate	2.06
Camphor	2.02	Benzil	2.08
Diphenylamine	2.06	Mercuric cyanide	2.09

the molecular elevation constant of methyl acetate have been obtained in dilute solution with the solutes mentioned; these may be compared with the figure 2.06 calculated from the boiling point and heat of evaporation

* It is a matter of experience (p. 712) that for a dilute solution Raoult's law is almost invariably applicable to the solvent; it is, therefore, usually a sufficient condition to state merely that the solution must be dilute for (25) to hold.

of the solvent. Similar results have been reported for other solvents, and the agreement between the theoretical and mean observed values for K_b is generally good; some of the relevant figures are recorded in Table 104.

TABLE 104. MOLAL CONSTANTS FOR ELEVATION OF BOILING POINT

Solvent	K_b		Solvent	K_b	
	Calc.	Obs.		Calc.	Obs.
Water	0.52	0.52	Methyl alcohol	0.83	0.8
Acetone	1.73	1.7	Benzene	2.60	2.7
Carbon tetrachloride	5.02	5.0	Chloroform	3.85	3.7
Ethyl alcohol	1.19	1.2	Ethyl ether	2.16	2.1

Since the boiling point of a solvent varies with pressure so also will the elevation constant be affected, both because of the change in T_0 and the accompanying change in L_v . In general the effect of these factors is within the limits of the experimental error, the value of K_b decreasing by approximately 0.03 per cent for a 10 mm. increase of pressure in the vicinity of 1 atm.

For concentrated solutions (25) can no longer be expected to hold; the reasons for this are considered more fully on p. 645 in connection with the depression of the freezing point, and the arguments apply equally to the elevation of boiling point.

Determination of Molecular Weights.—From the results quoted above it is apparent that, under the conditions specified, (24) and (25) may be regarded as reliable, and re-writing them in the form

$$M_2 = K_b \frac{1000w_2}{\Delta T_b w_1}, \quad (26)$$

they can be used for the study of molecular weights in solution. The experimental procedure is often described as the **ebullioscopic method** for the determination of molecular weights, and the whole subject is referred to as **ebullioscopy**. The principle of the method is that a definite weight w_2 of solute, whose molecular weight is to be found, is dissolved in a known weight w_1 of solvent and the elevation of boiling point ΔT_b is measured; since the molecular elevation constant for the solvent is known, M_2 can be calculated by (26). To avoid the possibility of error due to the volatility of the solute it is recommended that its boiling point should be at least 150° above that of the solvent. If these conditions cannot be met, and the solute is appreciably volatile at the boiling point of the solvent, it is possible to apply a correction (see footnote, p. 650).

Experimental Methods: (i) **The Beckmann Method.**—In the course of his work on solutions E. Beckmann (1888) developed the thermometer, bearing his name, which is still employed in this and other fields of measurement. It was based on a design by H. Walferdin (1840) and contained a reservoir at the top whereby it was possible to vary the amount of mercury in the bulb, so that the thermometer could be used over a wide range of temperature, from -39° to $+250^\circ$ c. It is obvious that such a thermometer cannot give actual temperatures, but for the measurement of the rise of boiling point differences only

are required, and these can be read with an accuracy of 0.001° . Strictly speaking a given length of stem in a Beckmann thermometer does not represent the same number of degrees over the whole temperature range; the discrepancy is relatively small and is neglected, except in work of high precision. A very simple differential thermometer has been devised by A. W. C. Menzies (1921); although it cannot be used over such a long temperature range as the Beckmann thermometer, it is more robust and is capable of measuring temperature differences with great accuracy.

The apparatus consists of a U-tube type of manometer containing water; one limb is short and wide, somewhat like a thermometer bulb, and the other is long and narrow, both being sealed up after evacuation leaving a free space above the liquid surface. Differences of temperature to which the bulb is exposed produce changes in the vapor pressure, so that the water rises to different levels in the narrow tube. The original paper gives a detailed table showing the equivalent in degrees of a change of 1 mm. in the water level; it decreases steadily from 0.0345° at 33°C . to 0.0025° at 102°C .

In the actual measurement of the rise of boiling point, two main sources of error are encountered; these are (i) fluctuations of temperature due to external air currents, and (ii) superheating of the liquid, that is the temperature may rise above the true boiling point. The various forms of apparatus used from time to time have been designed to overcome these difficulties. In the Beckmann apparatus, *which is now of historical interest only*, the tube A (Fig. 137) containing the boiling liquid has a thick platinum wire fused through the bottom so that the heat may be conducted through it into the solution and bubbles be formed at its upper end; the danger of superheating is thus diminished. Further, glass beads, garnets or, best of all, small platinum tetrahedra are placed in the

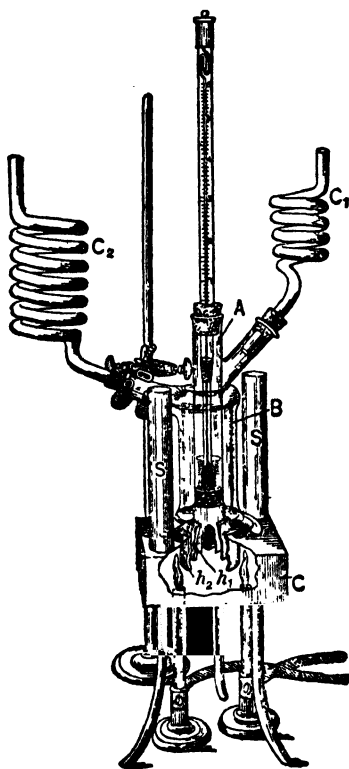


FIG. 137. Beckmann's boiling point apparatus

tube to facilitate steady boiling, for in this way small bubbles of vapor form and there is a greater possibility for equilibrium to be attained between vapor and liquid. Superheating results only when this equilibrium does not exist. The jacket B, surrounding the tube A, is made either of glass, as in the figure, or of porcelain with mica windows; it contains some of the solvent which is kept at the boiling point and has the effect of reducing radiation from the inner tube. As an additional precaution against superheating an indirect method of heating is employed; the asbestos box C has two asbestos rings h_1 and h_2 , and the gas burners are placed outside them; the tube A is protected from the direct action of the flames, while the hot gases pass up through the asbestos chimneys

S, S. Reflux condensers are attached to both *A* and *B* to minimize loss of liquid, and the whole apparatus is surrounded by a shield against draughts. A weighed amount, about 20 g., of solvent is introduced into *A* and its boiling temperature read on the Beckmann thermometer, when it has become steady. A weighed pellet of the solute under investigation is then added and the boiling point re-determined; the difference between the two temperatures gives the rise of boiling point. The molecular weight of the solute can then be determined by (26), the elevation constant K_b being obtained from tables.

Since 1888 many different types of apparatus have been described with the object either of simplifying the technique or of overcoming the danger of superheating. The use of electrical heating in the latter connection was proposed by S. L. Bigelow (1899) and employed by E. Beckmann (1908), J. H. Mathews (1911), and others; the heating current is carried by a fine platinum wire, about 20 cm. long and 0.4 mm., or less, diameter. Bubbles of vapor rise from the wire and there is little superheating; better results are said to be obtained with a wire of the thickness quoted above, employing a relatively high current (J. H. Mathews) than with a finer wire (0.1 mm.) and a lower current (E. Beckmann). The E.M.F. across the heating wire should be insufficient to permit electrolysis to occur.⁷

(ii) **The Landsberger Method.**—In 1892 J. Sakurai described a procedure for heating the solution to its boiling point by passing in the vapor from the boiling solvent. As the vapor condenses it gives up its latent heat to raise the temperature of the solution. In this method superheating is theoretically impossible, since when the boiling point is attained the liquid is in equilibrium with vapor at atmospheric pressure and no more vapor will condense, except in so far as it is necessary to supply heat to compensate for loss by radiation. The above principle was employed by W. Landsberger (1898) in an improved apparatus, and a further advance was made almost immediately afterwards by J. Walker and J. S. Lumsden (1898). The apparatus of these workers, still frequently employed, especially in teaching laboratories, is shown in Fig. 138; it differs from that of Landsberger only in the graduations on the tube *B* and the introduction of the bulb near the top to prevent loss of liquid by splashing. A small amount of solvent is placed in *B* and solvent in *A* is boiled so that the vapor passes into *B* where it emerges in small bubbles through a number of holes at the end of the tube. The temperature of the liquid rises and when it attains a constant value it is read on the thermometer; any excess of vapor passes out through a small hole on the right side of *F* near the top and then on to the condenser *D*. A weighed amount of solute is added to *B*, and the solution raised to its boiling point as before; when the temperature appears to be steady it is read, the passage of vapor is stopped and the volume of solution, as shown on the graduated tube *B*, is immediately recorded. The experiment may be repeated with further amounts of solute. A thermometer graduated in tenths of a degree, and capable of being read to 0.01° , is generally used in this work; although superheating is avoided, the temperature does not remain absolutely steady, because the condensation of vapor, necessary to compensate for heat losses, causes a slow dilution of the solution. To calculate the molecular weight of the solute, (26) is written in the form

$$M_2 = K_b \frac{1000w_2}{\Delta T_{\infty} \nu_1 \rho_1} = K'_b \frac{1000w_2}{\Delta T_{\infty} \nu_1}, \quad (27)$$

where $v_1\rho_1$ replaces w_1 , and ρ_1 , the density of the solution, or approximately that of the solvent, at its boiling point is included in K'_e ; v_1 is the measured volume of the solution. Values of K'_e , equal to K_e/ρ_1 , have been determined for several solvents, and some of the results are given in Table 105.

TABLE 105. ELEVATION CONSTANTS FOR THE LANDSBERGER METHOD

Solvent	K'_e	Solvent	K'_e
Water	0.54	Benzene	3.27
Acetone	2.23	Chloroform	2.60
Ethyl alcohol	1.58	Ethyl ether	3.03

Various modifications of the Landsberger apparatus have been described: in the one proposed by H. N. McCoy (1900), and subsequently modified by E. Beckmann (1902) and others, the tube for leading in the vapor is sealed into

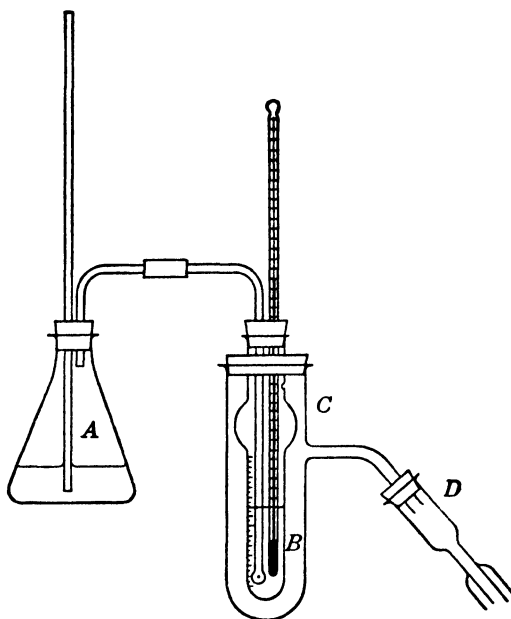


Fig. 138. Elevation of boiling point (Landsberger)

the side of tube B. The vessel for producing the vapor from the boiling solvent surrounds this tube and a separate flask is unnecessary; this part of the apparatus resembles the Menzies vapor pressure apparatus (Fig. 134). Loss of heat by radiation, and consequently condensation, is reduced to a minimum.⁸

(iii) **The Cottrell Method.**—An important advance in the technique of measuring boiling points of solutions was made by F. G. Cottrell (1919); he pointed out that even if there is no real superheating, the boiling points of most common solvents are approximately 0.1° higher at a depth of 3 to 4 cm. than at the surface on account of the hydrostatic pressure of the column of liquid. As a consequence, no matter how thoroughly equilibrium is attained, between

the liquid and bubbles of vapor, the thermometer really measures an approximate mean boiling temperature which varies throughout the depth of the liquid. In Cottrell's method the thermometer is placed in the vapor *above* the surface of the liquid and the boiling solution is made to pump itself continuously over the bulb of the thermometer. A thin layer of solution, which readily comes into equilibrium with vapor at atmospheric pressure, thus covers the bulb and so the thermometer should read the true boiling point, without errors due to superheating or hydrostatic pressure. The "pump" originally employed by Cottrell consisted of a piece of glass tubing, opened out at its lower end to a funnel shape; the tube was partly in the boiling liquid and partly out of it, the upper end reaching above the level of the thermometer bulb. When the liquid is boiled a stream of bubbles with entrapped liquid rises in the pump tube and is poured over the bulb which soon registers a steady temperature. The apparatus was improved by E. Washburn and J. W. Read (1919), who used a pump tube with two arms, and by J. F. Spencer (1921) who introduced a third arm and made the whole apparatus less fragile. A simple form is shown in Fig. 139; the tube *A* contains the boiling liquid, solvent or solution; a thick platinum wire *B* sealed into the bottom facilitates uniform ebullition. The side tube *C* leads to a condenser, and the sheath *D* prevents the cold condensate from reaching the thermometer *E*. The pump *F* is depicted with two arms, and when the liquid in *A* is boiled, by means of a small flame, bubbles and liquid trapped by the funnel are poured over the bulb. Electrical heating, as described above, may, with advantage, replace the gas burner (J. N. Pearce, 1926). It has been shown by W. D. Bancroft and H. L. Davis (1929) that the simple Cottrell method does not eliminate superheating entirely, and they devised an improved form of apparatus which reduced the extent of this error. A very accurate boiling-point apparatus, employing the Cottrell principle, was designed by B. Saxton and R. P. Smith (1932) and improved by R. P. Smith (1939), so that it could be used at pressures from 150 to 760 mm. The difference in temperature between the pure solvent and the solution, giving the rise of boiling point, is measured by a thermocouple with an accuracy of 0.0002° .⁹

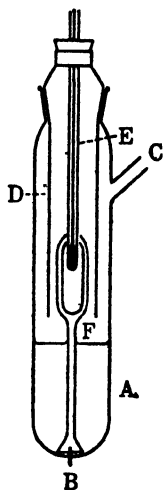


FIG. 139. Elevation of boiling point (Cottrell)

(iv) **General Improvements.**—When accurate ebullioscopic measurements are being made it is necessary to maintain the external pressure constant, for any change in the barometric pressure during the course of an experiment may introduce an appreciable error. The difficulty can be avoided by connecting the outlet of the boiling vessel to a manostat, i.e., an automatic device for maintaining constant pressure (E. Beckmann, 1912; J. H. Matthews, *et al.*, 1925). An alternative scheme was employed by E. Washburn and J. W. Read (1919); two similar pieces of boiling apparatus were used, one containing pure solvent and the other the solution, and the differences of temperature measured with two thermometers which had been compared with one another. The two temperatures necessary to determine the elevation of boiling point were thus observed simultaneously and hence at the same pressure. These authors also suggested that the temperature difference between boiling solvent

and solution could be determined directly by a thermocouple; this method was actually used by J. N. Pearce (1926), although it is of interest to record that a differential method, with two platinum resistance thermometers, had been employed for the same purpose by H. T. Barnes, *et al.*, in 1905. Of other refinements in ebullioscopic measurements, mention may be made of the introduction of a side-tube into the boiling vessel, whereby a sample of solution could be removed for analysis immediately its boiling point had been recorded (Washburn and Read), and of the use of a Dewar vessel (J. H. Mathews, 1911), or of a jacket containing the vapor of the boiling solvent (B. Saxton, 1932), both in conjunction with electrical heating, to diminish the loss of heat by radiation.

DEPRESSION OF THE FREEZING POINT

The Freezing Points of Solutions.—Another result of the lowering of the vapor pressure of a solution by a nonvolatile solute is that the freezing point is lower than that of the solvent. The fact that a dissolved substance depresses the freezing point of water had been known for many years, and in 1771 R. Watson observed that the time taken for a salt solution to freeze, and hence presumably the lowering of the freezing point, was proportional to its concentration. The same general conclusion was reached by C. Blagden (1788) and F. Rudorff (1861), neither of whom was apparently aware of previous work. Priority is generally, but probably erroneously, given to Blagden, and hence the proportionality between the lowering of the freezing point and the concentration of the solution has been called **Blagden's law**. In the years 1871–72 L. C. de Coppet showed that equimolecular amounts of different solutes when dissolved in the same quantity of water often gave the same depression of the freezing point; this work, like that of previous investigators, was limited, however, by the fact that it was carried out with aqueous solutions of salts. A significant advance was made by F. M. Raoult (1878–86); he employed both aqueous solutions of acids, bases and salts, and non-aqueous solutions of organic compounds, and as a result of his extensive work he established the fact that equimolar solutions of different substances of a similar type * lowered the freezing point of a given solvent to the same extent. In 1888 E. Beckmann turned his attention to the subject and obtained results of a much higher order of accuracy than those of previous workers; his work may be regarded as marking the beginning of the modern era. Further advances were made subsequently in the field of **cryoscopy**, as the study of the freezing points of solutions is called, so that such measurements can now be carried out with great precision.

Freezing Point and Vapor Pressure.—The connection between the lowering of the vapor pressure and the depression of the freezing point was realized by C. M. Guldberg (1870), F. M. Raoult (1878), and F. Kolářček (1882), although the first satisfactory theoretical treatment was

* Raoult found that the freezing point lowerings produced by equimolar concentrations in any solvent approached two mean values. It is now known that one of these is correct, whereas the other was due to the use of solutes which were associated, e.g., hydroxylic compounds in benzene solution, or dissociated, e.g., electrolytes in water.

not made until 1886 by J. H. van't Hoff. The relationship may be shown in a simple manner by means of the vapor pressure curves of a solvent, e.g., water, a solution, and the solid, e.g., ice, which separates on freezing (Fig. 140). The freezing point of the solvent is at A , temperature T_0 , where the vapor pressure curves for solvent and ice meet, since the liquid solvent and solid are in equilibrium at this point; similarly, the freezing point of the solution, at which it is in equilibrium with *pure* solid solvent, is at B , temperature T , so that $T_0 - T = \Delta T_f$, is the lowering of the freezing point.* The distance AC is equivalent to $p^0 - p$ at the

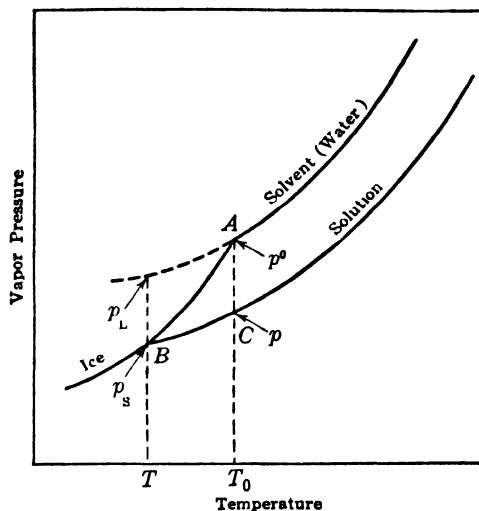


FIG. 140. Depression of freezing point

freezing point of the solvent, and since p^0 at this temperature is a constant, it follows, as on p. 634, that AC is proportional to the relative lowering of vapor pressure. If, as before, the vapor pressure curves for a number of dilute solutions are almost parallel in the vicinity of the freezing point, then BC/AC will be constant; for these solutions, therefore, the depression of the freezing point will be proportional to the relative vapor pressure lowering and hence to the mole fraction of the solute. It follows then for a dilute solution (cf. p. 634) that

$$\Delta T_f = k_f \frac{w_2}{M_2} \cdot \frac{M_1}{w_1}, \quad (28)$$

and so the lowering of the freezing point, like the rise of boiling point, should be proportional to the molar concentration, in agreement with

* Strictly speaking A and B are the freezing points at the respective equilibrium vapor pressures of solvent and solution (see p. 466), and are not those measured at atmospheric pressure. For a dilute solution, however, the depression $T_0 - T$ will be very little different from that at 1 atm. pressure

the experimental results already mentioned, and hence it is also a colligative property.

A more exact derivation of (28) is based on the use of the Clapeyron-Clausius equation. Assuming the vapor to obey the ideal gas laws, and its volume to be large in comparison with that of the solid or liquid, it follows that for the equilibrium between solid solvent (ice) and vapor, at constant external pressure,

$$\frac{d \ln p_s}{dT} = \frac{L_s}{RT^2}, \quad (29)$$

where L_s is the molar heat of sublimation (cf. p. 459). Similarly, for the equilibrium between liquid and vapor the corresponding equation is

$$\frac{d \ln p_L}{dT} = \frac{L_e}{RT^2}, \quad (30)$$

where L_e is the molar heat of evaporation of the liquid. If this equation may be assumed to hold for the supercooled liquid, then (29) and (30) will be applicable at the same temperature, e.g., T in Fig. 140, and hence

$$\frac{d \ln (p_s/p_L)}{dT} = \frac{L_s - L_e}{RT^2} = \frac{L_f}{RT^2}, \quad (31)$$

where L_f , the molar heat of fusion, is equal to $L_s - L_e$ (see p. 462). At the freezing point of the solution the vapor pressure of the solid (ice) must be equal to that of the solution (p_1) with which it is in equilibrium (see point B in Fig. 140); hence at the freezing point of the solution, (31) becomes

$$\frac{d \ln (p_1/p_L)}{dT} = \frac{L_f}{RT^2}. \quad (32)$$

Since p_1 is the vapor pressure of the solution and p_L that of the (supercooled) liquid at the same temperature, p_1/p_L is equal to x_1 if Raoult's law is applicable, so that

$$\frac{d \ln x_1}{dT} = \frac{L_f}{RT^2}. \quad (33)$$

When x_1 is unity, i.e., pure solvent, the freezing point is T_0 , and hence integration of (33) between T and T_0 , assuming L_f to remain constant, gives

$$\ln x_1 = -\frac{L_f}{R} \left(\frac{1}{T} - \frac{1}{T_0} \right) \approx -\frac{L_f}{R} \cdot \frac{\Delta T_f}{T_0^2}, \quad (34)$$

where ΔT_f is substituted for $T_0 - T$, and $T_0 T$ is put equal to T_0^2 ; this is justifiable if the solution is dilute. Since $x_1 + x_2 = 1$, $\ln x_1$ is equal to $\ln (1 - x_2)$, and for a dilute solution, i.e., when x_2 is very small, $\ln (1 - x_2)$ is approximately equal to $-x_2$, so that

$$\Delta T_f = \frac{RT_0^2}{L_f} x_2. \quad (35)$$

Finally, making the same assumptions as on p. 635,

$$\Delta T_f = \frac{RT_0^2}{1000l_f} \cdot \frac{1000w_2}{M_2w_1} \quad (36)$$

$$= K_f m, \quad (37)$$

where K_f , the molal depression constant of freezing point, or the cryoscopic constant, for the given solvent, is defined by

$$K_f = \frac{RT_0^2}{1000l_f}, \quad (38)$$

where l_f is the latent heat of fusion per gram, and m is the molality of the solution. The physical significance of K_f is seen to be the depression of the freezing point of a solution of unit molality, if the value were proportional to that for a dilute solution (cf. p. 636).

It will be evident from the deduction given above that (37) can only hold for very dilute solutions,* and the data in Table 106, for the freezing

TABLE 106. FREEZING POINT DEPRESSION OF SOLUTIONS OF CARBON TETRACHLORIDE IN BENZENE

m	ΔT_f	$\Delta T_f/m$	m	ΔT_f	$\Delta T_f/m$
0.1184	0.603°	5.093	0.5630	2.801°	4.975
0.1738	0.883°	5.080	0.8166	4.005°	4.904
0.3499	1.761°	5.031	1.166	5.620°	4.818

points of solutions of carbon tetrachloride (C. R. Bury and H. O. Jenkins, 1934), are of interest in this connection. The values of $\Delta T_f/m$, which by (37) should be equal to K_f , diminish somewhat with increasing concentration of the solution. If the results are extrapolated to infinite dilution K_f is found to be 5.11, in excellent agreement with the figure 5.12 obtained from (38), taking the freezing point of benzene (T_0) to be 278.6° K. and the latent heat of fusion (l_f) 30.12 cal. per g. There are three important reasons for the steady decrease in the values of $\Delta T_f/m$ in Table 106, which should be constant according to (37). First, the latent heat of the solid solvent is not independent of temperature, as was assumed in the integration of (33); an allowance could be made for this variation, if necessary for accurate work, by (85), p. 462. Second, it is only for relatively dilute solutions that $\ln x_1$ can be put equal to $-x_2$; and third, the replacement of T_0T by T_0^2 is only justifiable when the depression of the freezing point is small. It should also be noted that the deduction of (36) from (35) implies that the number of moles of solute is negligible. These sources of error are all inappreciable for dilute solutions, but have a marked influence as the concentration is increased.

For the determination of molecular weights by the freezing point method it is useful to know depression constants for a few common solvents; some of these which are applicable if the solutions are not too concentrated are given in Table 197.

* For relatively concentrated solutions, much better results are obtained by (34), or even (35).

TABLE 107. MOLAL FREEZING POINT DEPRESSION CONSTANTS

Solvent	K_f	Solvent	K_f
Water	1.86	Naphthalene	7.0
Acetic acid	3.90	Bromoform	14.3
Benzene	5.10	Cyclohexane	20.2
Nitrobenzene	6.9	Camphor	40.0

Determination of Molecular Weights.—Because of the accuracy with which freezing point measurements can be made, the method is frequently employed for the determination of molecular weights; for this purpose (36) and (37) are written in the form

$$M_2 = K_f \frac{1000w_2}{\Delta T_f w_1} \quad (39)$$

There are two important points to remember in the application of this equation: first, that the solutions must not be too concentrated, and second, that on cooling the solid which separates must be pure solvent. In some instances, e.g., a solution of iodine in benzene, a solid solution is deposited on cooling; in such cases the deductions given above do not hold and another treatment must be adopted, as shown below (p. 649).

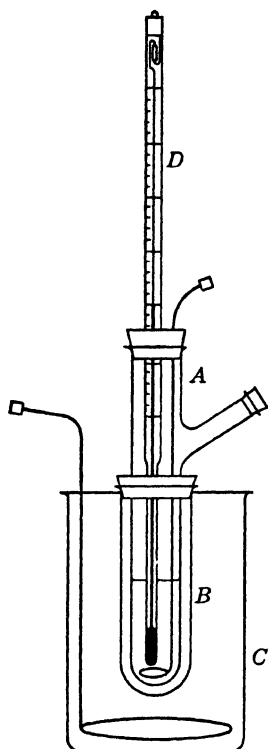


FIG. 141. Measurement of freezing point (Beckmann)

Experimental. (i) The Beckmann Method.—

The first accurate measurements of freezing point depressions were made by E. Beckmann (1888), using the thermometer (p. 637) specially devised for this purpose. His form of apparatus is still used in general laboratory practice, although for precision work improved methods are employed. A known weight, e.g., 20 g., of pure solvent is placed in the tube A (Fig. 141), with a Beckmann thermometer D and a stirrer; this tube is fitted inside another (B), leaving an air space, and the whole is immersed in the covered vessel C containing a freezing mixture not more than 5° below the freezing point of the solvent. The temperature of the liquid in A is allowed to fall until it is about 0.5° below its normal freezing point; that is to say, it is supercooled without solid separating. Vigorous stirring then causes crystallization to commence, and the temperature rises to the freezing point, which is then recorded. In some forms of apparatus the stirring is done mechanically by an electromagnetic make and break device (E. Beckmann, 1903). The tube A is next removed, warmed slightly to melt the solid and a weighed amount of solute is added and allowed to dissolve; the freezing point of the solution is determined by the same method as for the solvent, and the difference gives the depression of the freezing point. Sufficient data are now available

for the molecular weight of the solute to be evaluated, provided the depression constant of the solvent is known. Differential methods permit the lowering of the freezing point to be measured directly; one vessel contains the pure solvent and the other the solution, and the difference of temperature can be determined by a thermocouple (H. Hausrath, 1902) or by two platinum resistance thermometers (T. G. Bedford, 1910). The former procedure has been largely used in more recent work.

Sources of Error.—When the freezing point of a solution is being determined the quantity of solid solvent which has separated must be small, otherwise the concentration of the solution will be changed appreciably and the freezing point will not be that of the solution as made up. The magnitude of the error can be reduced by supercooling to a small extent only, and if freezing does not commence when the liquid is stirred, it may be encouraged by “seeding” or “inoculation,” that is, by the addition of a minute crystal of solid, e.g., ice. An important source of error in accurate freezing point measurement was considered by W. Nernst and R. Abegg (1894): equilibrium between liquid and solid phases is only attained relatively slowly, so that the apparently constant maximum temperature recorded by the thermometer is not necessarily the true freezing point. The actual reading depends on two opposing factors, viz., loss of heat by the contents of the tube *A* to the surrounding freezing mixture in *C*, and liberation of latent heat in the tube as a result of the separation of solid. The difference between the freezing point T_i recorded on the thermometer and the correct value T is given approximately by

$$T_i - T = k(T_c - T_i)/s, \quad (40)$$

where T_c , the “convergence” temperature, is the lowest temperature which the tube *A* would attain, under the experimental conditions, if no freezing occurred; it exceeds the temperature of the freezing mixture by a small, constant, amount. The constant k is related inversely to the heat insulation of the tube, while s , which is a measure of the amount of ice separated at equilibrium, may be taken as roughly equal to the extent of supercooling. One way of minimizing the discrepancy $T_i - T$ would be to increase the supercooling, but this would introduce an error due to a change of concentration of the solution. The only alternative is to keep $T_c - T_i$ small; that is to say, the temperature of the cooling bath should differ as little as possible from the freezing point being measured. The recorded temperature T_i is, of course, not an equilibrium temperature and so the thermometer reading should change until true equilibrium is attained; the rate of change is, however, so slow that it is not easily detected in practice.

(ii) **Equilibrium Methods.**—One way of overcoming some of the errors inherent in the Raoult-Beckmann supercooling procedure was proposed by W. Nernst and tried out by M. Roloff (1895). A quantity of aqueous solution was cooled until a relatively large amount of ice had separated and the whole system allowed time to come to equilibrium. The temperature was noted, and at the same time some of the solution, free from ice, was withdrawn and analyzed. As there was some danger that the solid separating on freezing may not have been pure ice, J. Walker (1903), T. W. Richards (1903), and others, suggested mixing the solution with ice frozen separately from the pure solvent and placing the mixture in a freezing bath, or in a Dewar vessel, for equilibrium to be attained. The composition of the solution was determined by analysis after the equilibrium temperature was reached. This principle has been used in most of the accurate

measurements of recent years, in combination with the differential method for the direct determination of the freezing point depression by means of a thermocouple; an accuracy of 0.00002° has been claimed in this work. In

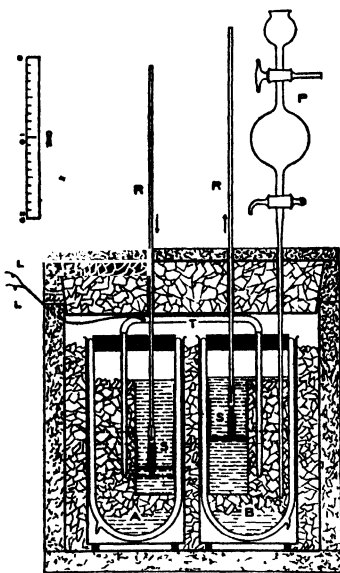


FIG. 142. Depression of freezing point

the apparatus (Fig. 142) of L. H. Adams (1915), two similar metal Dewar vessels, A and B, were packed round with crushed ice in a felt-coated box; each vessel contained about 400 g. of crushed ice and either 600 g. of water in A, or the same amount of solution in B. Thorough circulation of the liquids was brought about by the bucket-shaped stirrers SS, which were of the nature of pumps, of semicircular cross section. The pistons were perforated at the bottom and were fitted with movable discs to act as valves; these pistons were regularly raised and lowered by the glass rods RR. When rising, as in the vessel B, the valve closed the piston so that the enclosed liquid was carried up and poured over the ice; in falling, however, the movement of the valve, as shown in vessel A, allowed free passage of the liquid. The difference in temperature in the two vessels was measured by a thermocouple enclosed in T, the exit wires LL being connected to a potentiometer. When equilibrium was attained, as indicated by the constancy of the temperature reading, a quantity of the solution was withdrawn into the pipette P and analyzed. If

a steady difference of temperature is to be recorded the transfer of heat to or from the vessel B should be very small; it is with this objective that the Dewar vessels are used and are packed in ice.¹⁰

(iii) **Rast's Micro-Method.**—The molar depression constant of camphor is unusually high, viz., 40, and the use of this substance as a solvent for cryoscopic work was proposed by A. Jouniaux (1912); subsequently K. Rast (1922) described a method for the determination of molecular weights requiring small amounts only of substances soluble in molten camphor. A few milligrams of the experimental material are mixed with about ten times their weight of camphor and the whole melted; after solidification the mixture is finely ground and a quantity transferred to a capillary tube such as is commonly used for the approximate determination of melting points. The tube is then attached to an ordinary thermometer and the temperature raised in the usual manner, e.g., in a small sulfuric acid or paraffin bath. The point at which the last traces of camphor are about to be completely melted is recorded as the freezing temperature of the solution, for this is the point at which the liquid solution is in equilibrium with a small amount of solid solvent (cf. p. 750). The depression of the freezing point from that of pure camphor, measured in the same manner, is of the order of 10° or more, and so may be read with sufficient accuracy on a thermometer graduated in degrees. The results obtained are only approximate because such concentrated solutions are employed, but they are often of practical value. Determinations with camphor are somewhat inconvenient because of its high melting point (177°C); consequently, camphene, melting point

42.7° c., depression constant 35, has been suggested as an alternative. Other terpene derivatives, e.g., pinene dibromide, and fluorochloroethanes have been proposed as solvents for micro-determinations of molecular weight, for some have even higher depression constants than camphor.¹¹

Separation of Solid Solutions.—The equations deduced above for the lowering of freezing point are based on the assumption that solid solvent deposits in the pure state on cooling the solution; when both solute and solvent are present in the solid in the form of a solid solution (cf. p. 349), e.g., solution of iodine in benzene, the results are abnormal (J. H. van't Hoff, 1890). The obvious method for detecting the formation of solid solutions is to analyze the solid, but it is not always easy to obtain the latter free from solvent, and so the procedure of A. van Bijlert (1891) can be used. An inert reference substance is added to the solution which is then partially frozen; the mother liquor and wet solid are separated and both are analyzed. If the crystals contain no solute, then the whole of the latter associated with the wet solid will have come from the adhering solution, and the ratio of solute to reference substance will be the same as in the mother liquor. On the other hand, if the crystals are solid solutions and contain solute, the ratio of solute to reference substance will be greater in the wet solid than in the mother liquor.

The equation for the freezing point depression when solid solutions are formed may be derived in the following manner. The vapor pressure curves of pure solvent, liquid solution, pure solid and solid solution are shown in Fig. 143; *A* is the freezing point (T_0) of the solvent, and *B'* is that of the solution from which solid solution separates on freezing. The vapor pressures at various points are indicated in the diagram. For the solid solution-vapor equilibrium the integrated Clapeyron-Clausius equation is

$$\ln \frac{p'}{p''} = -\frac{L_s}{R} \left(\frac{1}{T'} - \frac{1}{T_0} \right), \quad (41)$$

where L_s is the molar heat of sublimation of solvent vapor from the solid solution; if the latter is dilute, then L_s may be taken as equal to the value for the pure solid solvent. Similarly, for the liquid solution-vapor equilibrium, it follows that

$$\ln \frac{p'}{p} = -\frac{L_e}{R} \left(\frac{1}{T'} - \frac{1}{T_0} \right), \quad (42)$$

where L_e is approximately equal to the molar heat of evaporation of the solvent, and subtraction of (42) from (41), making the same assumptions as previously (p. 644), gives

$$\ln \frac{p}{p''} = -\frac{L_1}{R} \frac{\Delta T'_f}{T_0^2}, \quad (43)$$

where $\Delta T'_f$ is equal to $T_0 - T'$, the observed depression of the freezing point. For a dilute solution obeying Raoult's law the quantity $-\ln p/p^0$ may be replaced by x_2 the mole fraction of solute in the liquid solution (see p. 644); if the solid solution may be assumed to behave ideally,

also obeying Raoult's law, then $-\ln p''/p^0$, where p'' is the vapor pressure of the solid solution and p^0 that of the pure solid at the same temperature, can be shown, in an exactly analogous manner, to be equal to

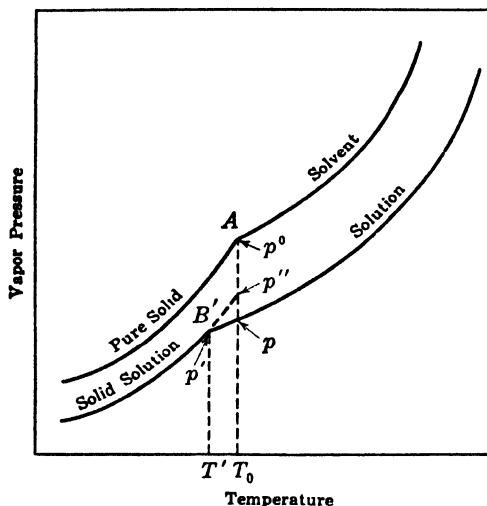


FIG. 143. Separation of solid solutions

x'_2 , the mole fraction of solute in the dilute solid solution. It follows, therefore, since

$$-\ln p/p^0 = x_2 \quad \text{and} \quad -\ln p''/p^0 = x'_2,$$

that

$$x_2 - x'_2 = -\ln \frac{p}{p''} = \frac{L_t}{R} \cdot \frac{\Delta T'_t}{T_0^2} \quad (44)$$

by (43); hence,

$$\Delta T'_t = \frac{RT_0^2}{L_t} (x_2 - x'_2) = \frac{RT_0^2}{L_t} x_2 \left(1 - \frac{x'_2}{x_2}\right), \quad (45)$$

which may be compared with (35), for which x'_2 is zero. The fraction x'_2/x_2 is the ratio of the mole fraction of the solute in the solid to that in the liquid solution, and consequently it may be replaced by the distribution coefficient (cf. p. 735), represented by k , so that

$$\Delta T'_t = \frac{RT_0^2}{L_t} x_2 (1 - k). \quad (46)$$

Finally, making the same simplifications as previously, it follows that

$$\Delta T'_t = K_t m (1 - k).^* \quad (47)$$

If the solute does not dissolve appreciably in the solid solvent, then k is

* An exactly analogous equation, viz., $\Delta T'_s = K_s m (1 - k)$, applies in connection with the rise of boiling point produced by an appreciably volatile solute; k is then the ratio of the mole fraction of solute in the vapor to that in the liquid phase.

almost zero and (47) reduces to the same form as (37). It will be noted that the presence of the factor $1 - k$ means that the freezing point depression when solid solution deposits must be less than for an equimolar amount of a solute which does not form solid solutions with the given solvent. In fact, if k is greater than unity, that is, the solute is more soluble in the solid solution than in the liquid, a *negative* depression, i.e., a *rise*, of freezing point may occur; the freezing point of the solution will then be higher than that of the solvent. Such cases are by no means uncommon, as will be seen later (p. 761); the vapor pressure of the solid solution in Fig. 143 is then lower and to the right of the one shown, and the point B' is to the right of A .

The following results obtained by E. Beckmann (1895-97) for solutions of iodine and thiophen in benzene, both of which separate solid solutions on freezing, indicate the accuracy of (47). K'_f is the apparent depression constant equal to $K_f(1 - k)$, from which k may be evaluated, and then compared with that obtained by direct experiment on the distribution of the solute between solid and liquid solutions.

	K'_f	$k(\text{calc.})$	$k(\text{obs.})$
Iodine in benzene	3.5	0.31	0.36
Thiophen in benzene	3.1	0.40	0.42

The agreement between the observed values of k and those calculated by (47) is reasonably good. One cause of deviation is the assumption that k is constant for all concentrations, whereas this is actually not the case since the solutions are not ideal.¹²

OSMOSIS AND OSMOTIC PRESSURE *

Semipermeable Membranes.—An important property of solutions, related to those already considered, is that of osmosis. The first recorded experiments of the phenomenon are those of the Abbé Nollet (1748) who found that when alcohol and water were separated by an animal bladder membrane, the water passed through into the alcohol, causing an increase of pressure, but the alcohol was not able to pass out into the water. This flow of water through animal membranes, found to occur with aqueous solutions in general, was studied by R. Dutrochet (1827-32) and by K. Vierordt (1848), the former of whom invented the terms "endosmosis" and "exosmosis," to describe the passage of water in opposite directions. Subsequently the prefixes were dropped and the word **osmosis** (Greek: *push*) was used to describe the spontaneous flow of water into a solution, or from a more dilute to a more concentrated solution, when separated from each other by a suitable membrane. The essential property of these membranes is that they allow the free passage of water but not of the dissolved substance, and J. H. van't Hoff (1886) suggested the adjective

* In reading this section the student is advised to clear his mind of any theories concerning osmotic pressure which he may hold as a result of an elementary study of the subject.

semipermeable to describe this property. The phenomenon of osmosis may be illustrated in a simple manner by tying an animal membrane, e.g., pig's bladder, over the end of an inverted thistle-tube which is then

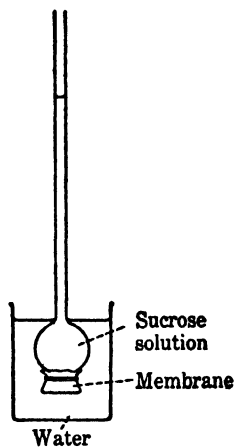


FIG. 144. Osmosis through animal membrane

partly filled with a concentrated solution of cane sugar and dipped into a beaker of water, as shown in Fig. 144. The water will rise in the tube until the excess hydrostatic pressure causes the water to flow outward at the same rate as the inward flow caused by osmosis. This particular arrangement does not lend itself, however, to the exact quantitative study of osmosis, as the animal membranes are not completely semipermeable; hence after a time sugar tends to pass out into the solvent and the level of liquid in the tube falls. Incidentally, the solution in the thistle-tube becomes considerably diluted by the entry of water.

Artificial Membranes.—As a consequence of T. Graham's discovery (1854) that substances in the colloidal state cannot pass through certain membranes which are permeable to water (p. 1232), M. Traube (1864), who was interested in the subject because of its importance in biology, was stimulated to search for artificial membranes able to prevent the passage of substances in true solution. Of the various materials examined, copper ferrocyanide, $\text{Cu}_2\text{Fe}(\text{CN})_6$, was found to be the most selective; it allowed the free passage of water but not of dissolved substances, as he demonstrated in the following manner. A small quantity of a copper salt solution was drawn into the end of a narrow tube and then gradually lowered into a solution of potassium ferrocyanide; a thin film of copper ferrocyanide immediately formed over the end of the tube, but it did not increase in thickness on long standing. This failure of the film to grow means that the copper and ferrocyanide ions cannot come into contact; that is to say, the copper ferrocyanide membrane does not allow these ions to pass through it, and once it is formed further interaction is stopped. The membrane also prevented the passage of other dissolved substances added to either the copper or ferrocyanide solution. At the present time copper ferrocyanide is still regarded as the best semipermeable membrane, being greatly superior to animal or vegetable membranes in this respect. Following on Traube's discovery important advances in the study of osmosis were made by W. F. P. Pfeffer (1877): first, by precipitating the copper ferrocyanide in the walls of a porous pot the membrane was enormously strengthened, and second, by the use of a closed manometer it was possible to measure the pressure due to osmosis without the solvent actually entering and diluting the solution under examination. This pressure is called the **osmotic pressure** of the solution, and it is best defined as *the excess pressure which must be applied to a*

*solution to prevent the passage into it of solvent when they are separated by a perfectly semipermeable membrane.** Actually no membrane is perfect in this respect, but the best modern membranes, described below, probably approach ideal behavior within the limits of normal experimental error.

Pfeffer's Experimental Method.—The apparatus used by Pfeffer is depicted in Fig. 145; the cell, containing the membrane in its pores, was attached by a wide T-shaped tube to a mercury manometer containing air in the closed limb. The membrane cells were prepared from porous pots made of fine-grained, unglazed porcelain, which were thoroughly cleaned and air completely removed from the pores by evacuation while the pot was immersed in water. The cell was next filled with 3 per cent copper sulfate solution and allowed to stand in some of the same solution for several hours. On removal it was quickly rinsed with distilled water, dried externally with filter paper and by standing in the air; a 3 per cent potassium ferrocyanide solution was then put into the interior and the cell again placed in the copper sulfate solution. The two electrolytes diffused from opposite directions into the pores where they interacted to precipitate a membrane of copper ferrocyanide. After one or two days the cell was closed up so as to allow an excess pressure to develop in the interior, because of the difference in osmotic pressure of the two solutions; this had the effect of pressing the copper ferrocyanide into the pores of the cell and so increasing the strength of the membrane. A further increase was attained by adding potassium nitrate to the interior of the cell and again closing it, so that an osmotic pressure of about 3 atm. was developed. A satisfactory membrane appeared when finished as a very fine, reddish-brown line in the white material of the pot; the semipermeability of the precipitate prevents it, of course, from growing in thickness. The T-tube and the manometer were attached to the pot, which was then filled with the experimental solution; it was finally closed with a rubber stopper and a glass tube passing through it was sealed off as shown, leaving no air in the cell. The whole apparatus was placed in a thermostat and the osmotic pressure determined from the manometer reading.¹³

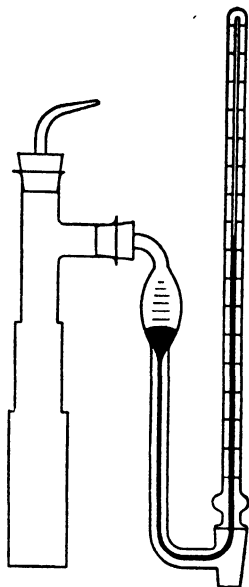


FIG. 145. Measurement of osmotic pressure (Pfeffer)

Concentration and Temperature.—Pfeffer made measurements with solutions of sucrose, dextrose, gum arabic, potassium nitrate and other substances; his results with sucrose have become of historical importance,

* The expression "osmotic pressure of a solution" has led to a great deal of confusion because it has sometimes been regarded as a pressure produced by the solution, and which causes the flow of solvent known as osmosis. It is much more satisfactory, however, to consider the osmotic pressure as being brought into existence only when the solution is separated from the solvent by a semipermeable membrane; the resulting osmosis then produces an excess pressure in the solution. Alternatively, if a pressure of the same magnitude is applied from the outside to the solution, the osmotic flow will not occur.

TABLE 108. OSMOTIC PRESSURE OF SUCROSE SOLUTIONS

Concentration (c) G per 100 g. Water	Osmotic Pressure (Π) Mm. of mercury	Π/c
1	535	535
2	1016	508
2.74	1518	554
4	2082	521
6	3075	513

for they were used by J. H. van't Hoff (1886) as the basis of a theory of solutions (p. 663). Some of Pfeffer's data, showing the influence of concentration on osmotic pressure, are recorded in Table 108; they were obtained at about 15° c. The approximate constancy of Π/c in the last column shows that the osmotic pressure of a solution is almost directly proportional to its concentration.

The effect of temperature is demonstrated by the results in Table 109, which Pfeffer obtained with a 1 per cent sucrose solution. It is evident that osmotic pressure increases with temperature, and the last column of the table shows the important relationship to which van't Hoff called attention, namely, the close proportionality between the osmotic pressure and the absolute temperature for a given solution.

TABLE 109. INFLUENCE OF TEMPERATURE ON OSMOTIC PRESSURE

Temperature	Absolute Temp. (T)	Osmotic Pressure (Π)	Π/T
6.8° c.	280.0° K.	505 mm.	1.80
13.7	286.9	525	1.83
22.0	295.2	548	1.85
32.0	305.2	544	1.79
36.0	309.2	567	1.83

Although Pfeffer's studies were a considerable advance over previous work, his results were by no means perfect, partly because of the incomplete semipermeability of the membranes used, and partly on account of inexact temperature control, as well as for other reasons. Nevertheless, little further progress was made until the beginning of the present century, when H. N. Morse and J. C. W. Frazer and their collaborators, and shortly afterward the Earl of Berkeley and E. G. J. Hartley, commenced a series of investigations which made such improvements in the difficult technique of osmotic pressure measurements as to give results of a high order of accuracy, especially for concentrated solutions. Before describing the methods, however, brief consideration will be given to the function of the semipermeable membrane.

Nature of the Semipermeable Membrane.—In the course of his work Pfeffer experimented with membranes of prussian blue and calcium phosphate; for a 1 per cent solution of sucrose the measured pressures were 387 mm. and 361 mm., compared with 535 mm. for copper ferrocyanide. These results seemed to imply that the osmotic pressure depended on the nature of the membrane used in its measurement, but it must be stressed that this is not the case.

There is only one osmotic pressure of a given solution at a definite temperature, but it can only be observed with a perfectly semipermeable membrane, and copper ferrocyanide is probably a close approximation. The other materials employed by Pfeffer were not truly semipermeable and allowed the passage, to some extent, of solute molecules, so that the pressures observed were less than the true osmotic pressure of the solution. Inefficient membranes of this type are said to be "leaky." The materials of which they are composed may be fundamentally incapable of functioning as a semipermeable membrane, or else the membranes may be weak physically so that they are broken down by excess pressure resulting from osmosis.

Provided a membrane is completely semipermeable, the osmotic pressure developed must be independent of the nature of the membrane; this is shown by the following theoretical consideration. Imagine a vessel divided into three sections by two fixed semipermeable membranes S_1 and S_2 (Fig. 146); the central portion, between the membranes, contains a solution, and pure solvent is placed outside. Suppose the osmotic pressure which can be produced at S_1 is P_1 , but at the membrane S_2 a smaller pressure P_2 only is attainable.

Solvent should flow through S_1 until an excess pressure P_1 exists in the solution; this pressure is, however, greater than that required to prevent osmosis at S_2 , so that solvent will pass out through this membrane in an attempt to reduce the excess pressure in the solution to P_2 . This decrease of pressure will disturb the equilibrium at S_1 where more solvent will flow in, only to pass out again at S_2 . There would be, consequently, a continuous flow of solvent from left to right, which could be utilized for the performance of work without the supply of energy from outside. In other words, in virtue of the two membranes setting up different osmotic pressures perpetual motion would be attainable. As this is contrary to the laws of thermodynamics, it is evident that osmotic pressure must be independent of the nature of the membrane. If one of the membranes, e.g., S_2 , is leaky, so that a lower pressure is attained thereat, there would be an initial flow of solvent, but it would soon cease; the solute would gradually pass out through S_2 and the concentration of the solution, and hence the osmotic pressure, would fall to zero. Some work could be obtained in this manner, equivalent to the energy of dilution.

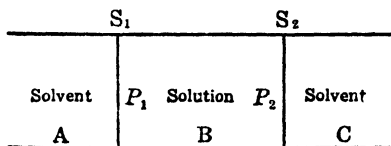


FIG. 146. Osmotic pressure and nature of membrane

The Cause of Semipermeability.—The earliest suggestions as to the origin of the semipermeable property of certain membranes was made by M. Traube (1867), who thought they acted as sieves, retaining the large molecules but allowing the passage of the smaller ones. The sieve mechanism may apply to some extent to the process of dialysis (p. 1232), but it is doubtful whether it accounts completely for the behavior of semipermeable membranes. It has been stated, for example, that membranes with pores many times larger in diameter than the molecules of solute are able to prevent the passage of the latter (S. L. Bigelow, 1907; E. F. Bartell, 1909, 1911). An alternative interpretation was proposed by M. L'Hermite (1855) which may be illustrated by the following experiment. Chloroform, water and ether are placed in a cylinder in the order given so as to form three layers. The ether being soluble in water

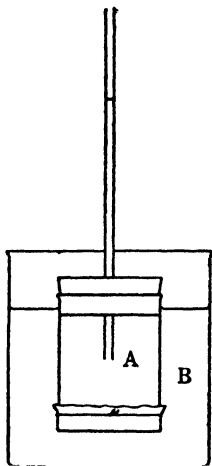


FIG. 147. Imitation of osmosis (Nernst)

tube, and placed in a beaker *B* containing moist ether. The level of the liquid in the narrow tube is seen to rise gradually, as in the osmosis experiment described on p. 652. The wet membrane behaves in an apparently semipermeable manner allowing ether, which is soluble in the water contained in the pores of the membrane, to pass from *B* to *A*, but the benzene, which is insoluble in water, cannot pass in the opposite direction.

From the experiments described it appears that semipermeability, in connection with osmosis, might be explained by supposing the solvent to be soluble in the membrane whereas the solute is insoluble; * it is not easy to see, however, what exactly is meant by solubility or insolubility in a substance like copper ferrocyanide. In recent years the view has been developed, based on J. von Liebig's suggestion (1862), that it is surface solubility, or adsorption (p. 1194), which is the factor determining semipermeability. Various qualitative and quantitative observations have been made in support of this idea; for example, F. Tinker (1916-17) found that copper ferrocyanide takes up water in preference to sucrose from an aqueous solution of this substance. In other words, water is preferentially adsorbed, and the sucrose exhibits what is known as "negative adsorption" (p. 1205). In general, a membrane is supposed to be permeable to the molecules it adsorbs positively, e.g., water, but impermeable to those, e.g., sucrose, which are negatively adsorbed.

It was considered at one time that the copper ferrocyanide constituting the membrane was amorphous in character, but later work involving diffraction of X-rays (J. F. Keggin and F. D. Miles, 1936) and of electrons (S. Fordham and J. T. Tyson, 1937) has shown that copper ferrocyanide, and other inorganic membranes, consist of a network of *crystalline* particles with an average diameter of about 1.5×10^{-6} cm. The dimensions of the pores in the membrane

* For an observation with gases, pointing in the same direction, see p. 665.

would probably be of the same order, and so they are undoubtedly too large for the membrane to act merely as a molecular sieve. Nevertheless, there is some connection between semipermeability and pore size, as is evident from the fact that a copper ferrocyanide membrane cannot entirely prevent the passage of solute molecules of low molecular weight * unless the pores of the cell are previously partially filled with an inert substance, e.g., magnesium silicate (see p. 658) (A. Grollman and J. C. W. Frazer, 1923). Tinker suggested that the pores of the membrane must be small enough for them to be under complete control of surface forces, which are responsible for adsorption, so that pore size is an important secondary factor.

If the particles constituting the membrane are covered with adsorbed molecules of solvent then, in spite of the network of small crystals, a continuity is established which permits the ready movement of solvent through the membrane. The molecules of solute, on the other hand, are unable to penetrate the complex network, especially as the preferential adsorption of solvent molecules on the surface makes the pores effectively smaller in diameter.

It must be noted that there may be no one explanation capable of covering all instances of semipermeability, and that there may be a danger in drawing conclusions from analogous phenomena. The surface of a solution consisting of a completely nonvolatile solute in a volatile solvent is a perfectly semipermeable membrane, since it allows free passage of solvent molecules, from liquid to vapor, but not of solute. The mode of operation of this semipermeable membrane may be quite different from that of a copper ferrocyanide, or similar, membrane, although H. L. Callendar (1908) has proposed a theory of osmosis (p. 667) which involves the idea of passage of vapor through capillaries. The semipermeable material is regarded as consisting of a large number of fine capillaries not wetted by the liquid solvent or solution, but through which molecules of vapor can pass. When pure solvent is placed on one side and solution on the other, distillation occurs through the capillaries from the region of higher (solvent) to lower (solution) vapor pressure, and hence osmosis results.¹⁴

Accurate Measurement of Osmotic Pressure.—Important advances in the measurement of osmotic pressure were made by H. N. Morse and J. C. W. Frazer and their collaborators (1901–23). Although a variety of gelatinous precipitates, e.g., uranyl, ferrous, zinc, cadmium and manganese ferrocyanides, ferric, uranyl and copper phosphates, and various cobaltcyanides were found to have semipermeable properties, none was superior to copper ferrocyanide. Improvements were, however, made in the method of preparing the semipermeable cell, as well as in the manometric device and in sealing the cell to this device. The first essential in the preparation of the cells is to obtain porous pots made from a fine-grained uniform clay without the addition of a "binder"; these are then washed, and air removed by endosmosis (p. 1220); the cell is filled with dilute copper sulfate solution and placed in a solution of potassium ferrocyanide. A copper electrode is inserted in the cell and one of platinum in the surrounding solution, and an electric current passed, the former electrode being the anode. The effect of the current is to drive the cupric and ferrocyanide ions, from inside and outside respectively, into the pores of the pot, where the copper ferrocyanide is precipitated. The resistance of the cell steadily increases and when a maximum is attained the current is stopped; the cell

* This accounts for the fact that nearly all reliable osmotic pressure measurements have been made with sugars, the molecules being relatively large.

is now washed, allowed to stand in distilled water for several days, and then again subjected to the electrolytic treatment. After several repetitions of this process, sometimes over a period of six months, cells capable of withstanding high pressures can be obtained. More recently, A. Grollman and J. C. W. Frazer (1923) have described a procedure whereby the pores of the pot may be partially filled with magnesium, or other, silicate prior to the deposition of the membrane; the latter may then be formed by ordinary diffusion, in a manner similar to that employed by Pfeffer (p. 653). In this way highly satisfactory cells may be prepared in a very short time.

In the earlier measurements the cell containing the solution was connected to a manometer and surrounded by the solution in the ordinary way (cf. Fig. 145), but later J. C. W. Frazer and R. T. Myrick (1916) reversed the arrangement, as shown in Fig. 148.

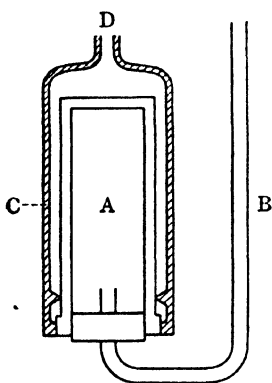


Fig. 148. Measurement of osmotic pressure (Frazer)

later (1921) by a water interferometer, utilizing the fact that the refractive index of water varies with the applied pressure.

A somewhat different principle for the measurement of osmotic pressure has been used by the Earl of Berkeley and E. G. J. Hartley (1906-09), following a suggestion made by G. Tammann (1892). Instead of observing the pressure developed as a result of entry of solvent into the solution through the semipermeable membrane, a gradually increasing external pressure was applied to the solution and the value determined at which the entry of solvent was just prevented. The semipermeable cell consisted of an open-ended porous tube, with the membrane deposited as near the outside as possible; this was achieved by soaking the tube in copper sulfate solution, drying superficially and then rotating in a solution of potassium ferrocyanide, with the ends plugged to prevent the formation of precipitate in the interior. The copper ferrocyanide membrane was then strengthened by the electrolytic method of Morse, described above. The tube (A, Fig. 149), containing the membrane, was filled with pure solvent and fixed, with liquid-tight joints, in a surrounding gunmetal vessel (B) containing the solution being studied. A hydrostatic pressure was applied to the solution through the tube C and the direction of flow of the solvent observed by the movement of the level of liquid in the capillary tube D. The pressure at which the movement reversed was taken as the osmotic pressure of

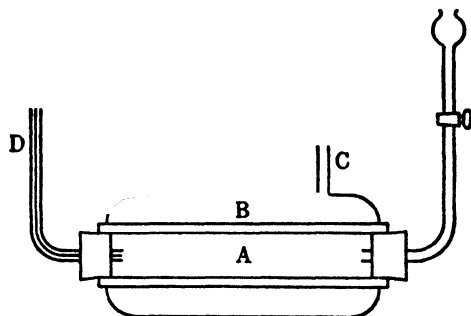


FIG. 149. Measurement of osmotic pressure (Berkeley and Hartley)

the solution for, in accordance with the definition (p. 652), this is the pressure which must be applied to the solution to prevent the entry of solvent through the semipermeable membrane. With the apparatus described Berkeley and Hartley measured osmotic pressures up to 150 atm.; it has the advantage of giving results in a very short time. The agreement between their results and those of Frazer and Myrick (1916) is shown in Table 110; the former, ob-

TABLE 110. OSMOTIC PRESSURE OF SUCROSE SOLUTIONS AT 30° C.

Sucrose per 1000 g. Water	Berkeley and Hartley	Frazer and Myrick
202 g.	15.48 atm.	15.59 atm.
370	29.72	29.78
569	48.81	47.88
820	74.94	73.06
1133	111.87	109.10
1430	148.46	148.80

tained at 0° c., have been recalculated to 30° c., on the assumption that the osmotic pressure is proportional to the absolute temperature. In view of the fact that some error is introduced by the temperature correction, the agreement is excellent, and may be taken as confirming the accuracy of both sets of measurements.

A dynamic method has also been described by Berkeley and Hartley (1909) and recommended especially for determining the osmotic pressures of dilute solutions, but it has not been employed to any extent. The actual rate of flow of solvent into a solution through a semipermeable membrane is measured and compared with the rate of flow through the same membrane under the influence of a mechanical pressure only. The pressure capable of producing a rate equal to that of the osmotic flow is taken to be equal to the osmotic pressure of the solution. It was found that, provided the *initial* rate of flow into solution was considered, the results were in good agreement with those given by the static method; in concentrated solutions, however, there were quite considerable deviations.

Many measurements of osmotic pressure, in both aqueous and nonaqueous solutions, have been made by other workers, but they are probably not of a very high order of accuracy, and need not be considered further. It may be mentioned, however, that osmotic pressure determinations, utilizing membranes of a simple character, have been performed on aqueous and nonaqueous solu-

tions of substances of high molecular weight. Further reference to these will be made later (p. 672).¹⁵

"Negative Pressure" Method for Measuring Osmotic Pressure.—A novel procedure for determining osmotic pressure was tried out by R. V. Townend (1928), at the suggestion of J. C. W. Frazer and W. A. Patrick, and later improved by F. T. Martin and L. H. Schultz (1931), by W. C. Eichelberger (1931) and by others; it makes use of the fact mentioned above that the surface of a solution is a semipermeable membrane. The method may be explained with the aid of Fig. 150; the essential part of the apparatus is the porous disc *A*, made from a mixture of special clay and powdered glass, the capillaries being large enough to prevent the passage of liquid through them, but so small that the maximum capillary rise is greater than the osmotic pressure to be measured. Below the disc in the tube *B* is solvent, and a tension can be placed on this by means of the mercury in *C*. The wide tube into which the disc is sealed is surrounded by a bulb containing the solution, and air is removed by a pump as shown. The rate of transfer of solvent from pure solvent to solution, or *vice versa*, is found by observing the movement of the junction between the solvent and mercury at *D*; this rate is varied by altering the tension on the solvent and the value found for which there is no movement in either direction. The difference of pressure between solvent and solution, as measured by the column of solvent in *B* and the mercury levels in *C*, is then equal to the osmotic pressure.* This method can only be used for low osmotic pressures, and the experimental technique is not easy. It has already been applied to the study of nonaqueous solutions, and it is probably in this direction that the method should prove most valuable,

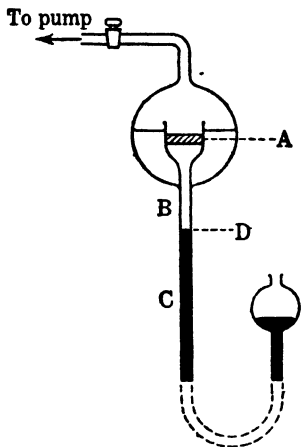


FIG. 150. Osmotic pressure by porous disc method

because of the difficulty of preparing artificial semipermeable membranes for such solutions.¹⁶

Comparison of Osmotic Pressure.—An approximate method for comparing osmotic pressures, involving the use of plant cells, was described by the botanist H. de Vries (1884); it is known as the **plasmolytic method**. A living plant cell consists, in general, of a relatively rigid cellulose wall which encloses the colloidal protoplasm; in the latter there are found the vacuoles of varying size, containing the cell sap consisting of a solution of salts, sugars and often coloring matter. The protoplasm lining the cell wall, that is, the cytoplasm, is bounded on its outer surface, nearest the wall, by the plasma membrane, and on its inner surface, enclosing the sap, by the vacuolar membrane (Fig. 151). Both membranes are selectively permeable, permitting the free passage of water molecules, but not of dissolved substances, such as sugars, of relatively high molecular weight. The cellulose wall of the cell is, however, permeable to both water and solute molecules. In a normal healthy cell water has entered through the wall, passed through the two membranes, and set up an

* One way of looking at the matter is that the amount whereby the pressure on the solvent must be decreased in order to reduce its vapor pressure to that of the solution is equal to the osmotic pressure (see p. 667).

excess pressure in the interior of the cell which causes the cytoplasm to be forced tightly against the walls; this condition is known as "turgor," and the cell is said to be "turgid." If such a cell is placed in water, or in a solution of osmotic pressure less than that of the cell sap, little change will be observed owing to the rigidity of the walls. On the other hand, if the cell is immersed in a solution having a higher osmotic pressure, water will pass outward from the

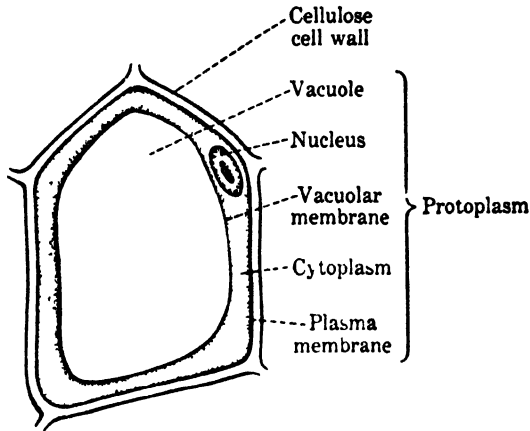


FIG. 151. Turgid plant cell

interior of the cell through the selectively permeable membranes, with the consequent loss of turgidity. The cytoplasm will then shrink and detach itself from the wall, as shown in Fig. 152; this phenomenon is called "plasmolysis." By taking a number of solutions of a given substance at different concentrations and observing the change in the cell, it is possible to find the solution which just fails to bring about plasmolysis; this solution has then the same osmotic pressure as the cell sap, and is said to be *isotonic* (Greek: *same tension*) with it. Working in the same manner with a variety of solutes, and a given type of cell, it is possible to prepare a series of isotonic solutions of different substances; such solutions have approximately the same osmotic pressure. If the cell membranes were completely semipermeable, these isotonic solutions would have exactly the same osmotic pressure, that is they would be iso-osmotic; this is, however, not the case since the membranes allow the passage of dissolved substances to some extent. De Vries expressed the results of his work in terms of the "isotonic coefficients," which were proportional to the reciprocals of the molar concentrations of the isotonic solutions of different solutes. One of the consequences of the studies of plasmolysis has been to show that, as a general rule, non-

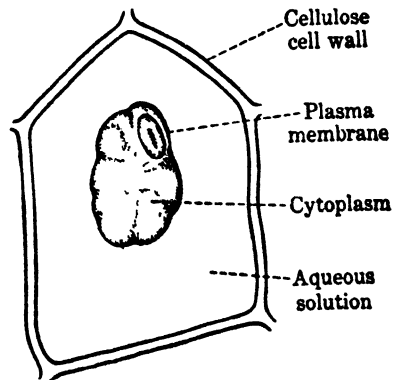


FIG. 152. Plasmolyzed plant cell

electrolytes have equal isotonic coefficients, so that equimolar concentrations have the same osmotic pressure. Solutions of electrolytes give different coefficients, since their behavior is abnormal (p. 888).

For the experimental work on plasmolysis almost any plant cells may be used, but those having strong cellulose walls and a colored cell-sap are most effective, e.g., cells from the leaf of *Tradescantia discolor* or from the root of the beet. Thin sections of the material, containing a considerable number of cells, are cut and placed on a slide and the changes resulting from the application of different solutions examined under a microscope. *Spyrogyra* and hairs from the stamens of *Tradescantia virginiana*, which consist of filaments one cell in width, have been found to be specially suitable for plasmolytic studies as it is unnecessary to cut sections.

An analogous method to that given above has been described by H. J. Hamburger (1886) for comparing osmotic pressures by the use of red blood corpuscles; these corpuscles are enclosed by a selectively permeable membrane, but unlike plant cells have no supporting cell wall so that they are relatively easily ruptured. If a small quantity of blood is added to water, or to a dilute aqueous solution, osmosis occurs and water passes through the outer membrane into the corpuscles; the latter thus become considerably distended and eventually burst, when the coloring matter, i.e., hemoglobin, is able to escape from the interior of the cells and give a red color to the solution in which they are suspended. When this occurs the corpuscles are said to have been "hemolyzed" or to have undergone "hemolysis." Suppose, on the other hand, the blood is added to a relatively concentrated aqueous solution; osmotic flow occurs in the opposite direction, that is, outwards from the interior of the corpuscles; these will thus shrink and become "crenated." After a short time the crenated particles settle to the bottom leaving a colorless supernatant liquid. By making experiments with solutions of different concentrations it is possible to find one just able to prevent the occurrence of hemolysis; such solutions of different substances are isotonic. The results obtained in this manner by Hamburger and others are very similar to those given by the plasmolytic method. It should be mentioned that the isotonic solutions found with the aid of blood corpuscles are not necessarily isotonic with the contents of the latter; the reason for this discrepancy is that sometimes the membranes are so strong that a considerable excess pressure is required to weaken them sufficiently to allow the hemoglobin to escape. A solution which just fails to cause hemolysis has thus, generally, a lower osmotic pressure than the serum within the corpuscles; the difference increases with the strength of the membrane.¹⁷

Mechanism of Osmotic Pressure: Dilute Solutions and the Gas Laws.

—While J. H. van't Hoff was studying the problem of equilibria in gases, his attention was drawn by H. de Vries to the work of Pfeffer on osmotic pressure (p. 653); this led to the publication in 1886 of the classical paper in which he pointed out the analogy between gases and solutions, and also applied for the first time thermodynamic methods to the colligative properties of solutions. It has been seen (p. 654) that Pfeffer's results indicated that the osmotic pressure at constant temperature is directly proportional to the concentration of the solution; van't Hoff showed that this was the analogue of Boyle's law for gases. Since the concentration c

in moles per liter is the reciprocal of the volume in liters containing 1 mole, it follows from Pfeffer's data that for a solution

$$\Pi/c = \text{constant}, \quad \therefore \Pi V = \text{constant}, \quad (48)$$

where Π is the osmotic pressure. Further, the proportionality between osmotic pressure and absolute temperature shows that a law equivalent to Gay Lussac's law for gases applies to solutions, so that

$$\Pi/T = \text{constant}, \quad (49)$$

and combining the two laws for solutions it is seen that

$$\Pi V = RT, \quad (50)$$

where R is a constant. By comparing the osmotic pressure of a cane sugar solution with the pressure exerted by hydrogen gas at the same temperature and concentration, van't Hoff showed that R for 1 mole of solute had almost the same value as the gas constant per mole. The results in Table 111 give osmotic pressures of a solution of sucrose containing

TABLE 111. COMPARISON OF OSMOTIC AND GAS PRESSURES

Temperature	Osmotic Pressure	Ideal Gas Pressure
6.8° C.	0.664 atm.	0.665 atm.
15.5	0.684	0.686
22.0	0.721	0.701
36.0	0.746	0.735

1 g., i.e., 1/342 mole, in 100.6 cc. of solution, observed by Pfeffer, compared with the pressure exerted by an ideal gas at the same concentration. The excellent agreement implies that the gas laws hold almost quantitatively for dilute solutions; that is to say, the analogue of Avogadro's law is also applicable. Equal numbers of molecules of different solutes dissolved in the same volume of solution give the same osmotic pressure, at the same temperature. This is in accord with the conclusion reached from the work of de Vries on isotonic solutions of nonelectrolytes (p. 661).

The correspondence of R in (50) with that in the gas law equation may be shown in another way. The osmotic pressure of a solution of sucrose containing 0.02922 mole per liter was found to be 0.655 atm. at 0° C.; expressing Π in atm. and V in liters per mole of solute, it follows that

$$R = \frac{\Pi V}{T} = \frac{0.655}{273 \times 0.02922} = 0.0815 \text{ liter-atm. deg.}^{-1} \text{ mole}^{-1}.$$

The value of R for solutions is thus very close to that of R usually accepted for a gas, namely 0.0821 liter-atm. deg.⁻¹ mole⁻¹, so that it is possible to write $\Pi V = RT$, where R may be taken as the gas constant.

The Bombardment Theory.—More recent accurate determinations of the osmotic pressures of dilute solutions have provided general support for van't Hoff's views, although the results show that the osmotic pres-

sure is usually a few per cent greater than the corresponding gas pressure; it will be seen later that for relatively concentrated solutions, however, having osmotic pressures greater than about 5 atm., i.e., concentrations greater than 0.2 M, the deviations are quite considerable. It is important to point out in this connection, as van't Hoff emphasized, that the gas law equation could only be expected to apply to solutions so dilute as to be comparable with ideal gases. The equation $\Pi V = RT$ has thus become known as the expression of van't Hoff's law for an ideal dilute solution; alternatively, it may be written $\Pi = RTc$, where c is the concentration in moles per liter. The van't Hoff equation implies that the osmotic pressure of a dilute solution is equal to the pressure which would be exerted by the same number of molecules of solute if they existed as a gas occupying a volume equal to that of the solution. The remarkable similarity between the behavior of gases and solutions led van't Hoff to suggest that gas pressure and osmotic pressure had the same fundamental origin. "In the former case," he wrote, "the pressure is due to the impacts of gaseous molecules on the walls of the containing vessel, and in the latter to the impacts of the molecules of dissolved

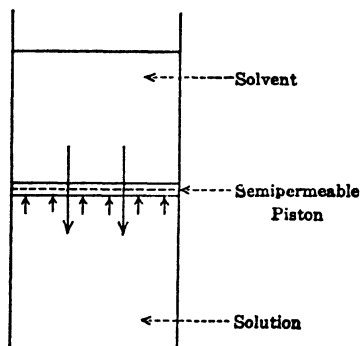


FIG. 153. Bombardment theory of osmotic pressure

substance on the semipermeable membrane." It may not be evident, at first sight, how impacts on the membrane by solute molecules can result in water being drawn into the solution, but that it is a necessary consequence may be readily shown. Imagine a cylinder (Fig. 153) containing solution and solvent separated by a tightly fitting, but movable, semipermeable piston; the molecules of solute bombard the piston from below without being able to penetrate it, and consequently the piston will move upward through the solvent. In other words, as a result of the bombardment by the solute, molecules of solvent will pass through into the solution.

If the piston is fixed this tendency for solvent to enter will still occur, and will become manifest if the solution possesses a free surface, as in the experiment on p. 652.

A further analogy between gases and solutions, which has been quoted as an argument for the solute bombardment theory, is shown by the observations first made by W. Ramsay (1894), and subsequently by others. When palladium is heated to about 300° C. it allows hydrogen gas to pass through it readily, but not nitrogen, and so this metal can be regarded as a semipermeable membrane for a solution of nitrogen in hydrogen. A mixture of nitrogen, at known partial pressure, and hydrogen was placed in a palladium bulb *A* attached to a manometer *B* (Fig. 154); the bulb was inserted in a vessel *C* through which hydrogen at atmospheric pressure was

streamed. On heating the bulb hydrogen passed through, and when equilibrium was attained the pressure inside the bulb was greater than atmospheric by an amount equal to the partial pressure of the nitrogen in the original mixture. This experiment is the exact parallel to that in which osmotic pressure is measured. The solution in a semipermeable vessel is surrounded by solvent and the latter then enters the solution setting up an excess pressure and, as with the nitrogen and hydrogen mixture, this may be identified with the partial pressure of the substance, i.e., the solute, to which the membrane is not permeable. Incidentally, this behavior of heated palladium, in allowing the passage of hydrogen only, suggests that in this case at least it is connected with solubility of the material in the membrane, since hydrogen is known to be soluble in heated palladium whereas nitrogen is not.¹⁸

Objections to the Bombardment Theory.

—The theory of van't Hoff concerning the origin of osmotic pressure, as the bombardment pressure of solute molecules, was at one time widely accepted, and is still to be found in textbooks, but in recent years there has been a growing opinion that it should be discarded because it has many serious objections. Consider the experiment on nitrogen and hydrogen described above: it is found that the hydrogen enters the palladium bulb until the pressures inside and outside are the same. When it is asked why the hydrogen enters the bulb, the answer is that it is at a lower partial pressure inside than outside, and not merely because there is nitrogen in the bulb exerting a pressure. In exactly the same way it may be said that the molecules of solvent enter the solution through a semipermeable membrane because their "partial pressure," or its equivalent, is less in the solution than in the solvent. It was assumed by van't Hoff that the impacts of the molecules of the solvent, being equal and opposite on both sides of the membrane, may be neglected; but this is by no means true. It is easily seen that there are more solvent molecules per unit volume, or per unit area, in the solvent than in the solution; consequently the impacts on the two sides of the membrane will not be the same. In fact, it might appear more logical to regard osmotic pressure as being due to the difference in bombardment pressure of the solvent molecules in the pure solvent and in the solution, rather than to the bombardment by solute molecules.

This aspect of the subject has been developed by F. Tinker (1916-17), who ascribed a definite "thermal pressure" to the solvent molecules; it is equal to the pressure they produce by bombardment in the interior of the liquid. It differs from the actual bombardment pressure at the walls, the two being related by an equation of the type employed by Dieterici (p. 296) in connection with gases.

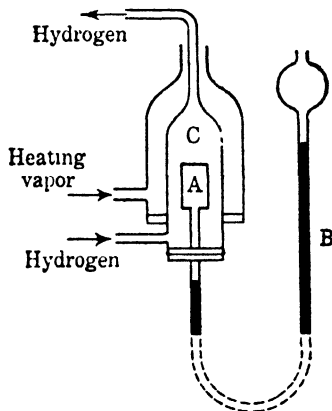


FIG. 151. Permeability of palladium to hydrogen (Ramsay)

The thermal pressure is supposed to have a smaller value in the solution than in the pure solvent, because of the smaller number of solvent molecules in a given volume. When the two liquids are separated by a semipermeable membrane, the difference in thermal pressure will result in a flow of solvent from the region of higher to that of lower pressure; that is, osmosis occurs. The osmotic pressure is nevertheless not equal merely to the difference in the two thermal pressures, and by a simple thermodynamic cycle it is shown that for an ideal solution

$$\Pi = \frac{RT}{\bar{V}_1} \ln \frac{P}{P'}, \quad (51)$$

where \bar{V}_1 is the partial molar volume of the solvent (see p. 239) and P and P' are the thermal pressures of the solvent molecules in the interiors of pure solvent and of solution, respectively. At high dilutions this equation becomes identical, as it should, with van't Hoff's law. A similar idea, in which the osmotic pressure is assumed to be determined by the difference in the pressure of solvent molecules on the two sides of the semipermeable membrane, has been worked out by G. Schay (1923), the van der Waals equation being used as a basis for the calculation of pressures. The conclusion reached is that only at infinite dilution could the osmotic pressure be equal to the bombardment pressure of the solute molecules.

A further objection to van't Hoff's theory is that even if osmotic pressure were due to impacts of solute, the pressure so obtained would not be the same as the pressure produced by the same number of molecules in the gas phase, since the free space in a solution is only a small fraction of the total volume. Some authors, however, have considered that the free space in solution is to be taken as all except that occupied by solute molecules, the volume of the solvent molecules being disregarded; it is doubtful if this view is justifiable. It has been argued that in Ramsay's experiment with nitrogen and hydrogen, described above, whatever may be the cause of the entry of hydrogen, the fact remains that at equilibrium the difference of pressure between the inside and outside is equal to the pressure of the nitrogen, and hence is due to the bombardment of the walls by molecules of the latter; by analogy, therefore, osmotic pressure would have to be ascribed to the impacts of solute molecules. The point is, however, that in a mixture of gases Dalton's law of partial pressures is obeyed (p. 302), at least approximately, so that the excess pressure *must be* equal to that of the nitrogen, but it does not necessarily follow that the same argument applies to osmotic pressure, except perhaps at infinite dilution. There is no doubt of the similarity between the equation $\Pi V = RT$ for a dilute solution and the characteristic equation for an ideal gas, but this identity gives no information whatever concerning the mechanism of osmotic pressure. It was shown by van't Hoff, and later by Lord Rayleigh (1897), and others, that when an ideal gas is dissolved in a nonvolatile solvent, it is a thermodynamic necessity that ΠV for the solution should be equal to PV for the gas; this result is quite independent of any mechanism of gas pressure or of osmotic

pressure. It is true that this deduction involves a solution of an unusual type from the point of view of the study of osmotic pressure, but the equation $\Pi V = RT$ can also be deduced thermodynamically, for a *very dilute ideal solution* of a nonvolatile solute, by an alternative method given below (p. 671). It is, therefore, entirely unjustifiable to draw conclusions from the analogy between the equations for a dilute solution and an ideal gas.

In addition to the thermal pressure theory, other interpretations of osmotic pressure involving properties of the solvent as modified by the presence of solute have been proposed. An attempt was made by G. Jäger (1891) to relate osmotic pressure to the difference in surface tension of solvent and solution, and similar views were expressed by B. Moore (1894) and I. Traube (1904), but there are so many objections to this theory that it need not be considered further, except to mention one point. Some solutions have surface tensions greater than that of water, but others have lower values; nevertheless osmosis always occurs in the same direction, i.e., from solvent to solution.

It was shown (p. 446), in connection with the study of vapor pressure, that the vapor pressure of a liquid may be increased by applying an external pressure; consequently by increasing the pressure on a solution its vapor pressure may be increased until it is equal to that of the pure solvent. The solution will then be in equilibrium with the solvent, and it is therefore thermodynamically necessary that the pressure exerted on the solution should equal the osmotic pressure. Similarly, reduction of the pressure on the pure solvent will diminish its vapor pressure and, as already recorded (p. 660), a method for the measurement of osmotic pressure has been based on the assumption that the latter is equal to the amount whereby the pressure on the solvent must be decreased to bring it into vapor equilibrium with the solution. These relationships between vapor pressure and osmotic pressure have been made the basis of a theoretical treatment. According to G. A. Hulett (1902) the osmotic pressure is to be regarded as a "negative pressure" applied to the solvent in order to decrease its vapor pressure to that of the solution, and A. Thiel (1937) has argued in favor of a similar point of view. H. L. Callendar (1908), on the other hand, considered that a positive excess pressure has to be applied to the solution so that its vapor pressure is increased; this interpretation of osmotic pressure was combined with a theory of the function of the semipermeable membrane (p. 657).¹⁹

Review of Theories.—There are, in general, three main views concerning the origin of osmotic pressure involving, respectively, solute bombardment, solvent bombardment, and vapor pressure effects, but it does not seem possible to decide between them. In fact, the different ideas may not be independent, for it is probably the motion of the solute molecules which causes the thermal pressure and the vapor pressure of the solvent to be less in solution than in the pure solvent. Nevertheless, it does not follow that the osmotic pressure must be *equal* to the pressure produced by impacts of solute molecules, and this theory is not so widely held now as it was at one time. It may be noted, however, as van't Hoff himself pointed out (1892), that the actual mechanism of osmotic pressure is not

of great importance, and even if the theory of solute bombardment is completely discarded van't Hoff's fundamental contribution to the subject of solutions still remains.

In the study of osmotic pressure the essential point to be borne in mind is that, for some reason connected with the presence of solute molecules, the partial free energy, chemical potential or activity of a solvent molecule is less in a solution than in the pure liquid; that is to say, the transfer of solvent from pure solvent to solution will result in a decrease of free energy. Such a transfer will, therefore, always tend to occur whenever solvent and solution are brought together, and the process will continue until equilibrium is attained. If the two liquids are placed side by side in a closed space the difference of free energy, or chemical potential, will manifest itself by distillation from solvent to solution; if two solutions of different concentration are used, the process will continue until they have the same vapor pressure (p. 632). Similarly, when solvent and solution are separated by a semipermeable membrane, solvent must pass into the solution until equilibrium is attained by the building up of an excess pressure. It follows, therefore, that the setting up of an osmotic pressure is the inevitable result of the introduction of a semipermeable membrane between a solvent and solution, on account of the difference of the chemical potentials of the solvent molecules. The vapor pressure of the solution, when equilibrium is attained under the influence of the osmotic pressure, must then, as a thermodynamic necessity, be equal to the normal value for the pure solvent.

Osmotic Pressure and Vapor Pressure.—From what has been said above it is clear that there must be a relationship, derivable thermodynamically and hence independent of mechanism, between the osmotic pressure of a solution and the lowering of vapor pressure. Suppose a solvent and a solution are separated, at constant temperature, by a semipermeable membrane. Let P_0 and P , respectively, be the pressures exerted on the two liquids necessary to maintain equilibrium; then the osmotic pressure Π of the solution is $P - P_0$. If μ_0 is the chemical potential of the pure solvent, and μ that in the given solution, both at pressure P_0 , then the two values will be different. When the equilibrium osmotic pressure is established the pressure on the solution will be increased to P , and the chemical potential will then be equal to that of the pure solvent. It follows, therefore, that

$$\mu_0 = \mu + \int_{P_0}^P \left(\frac{\partial \mu}{\partial P} \right)_T dP. \quad (52)$$

According to (169), p. 239, for a system of constant composition, $\partial \mu / \partial P$ at constant temperature is equal to the partial molar volume of the constituent under consideration, and so in the present case (52) becomes

$$\mu_0 = \mu + \int_{P_0}^P \bar{V}_1 dP, \quad (53)$$

where \bar{V}_1 is the partial molar volume of the solvent in the solution. If the partial vapor pressures of the solvent in equilibrium with the pure solvent and solution are p^0 and p , respectively, then the chemical potentials of the vapors are

$$\mu_{0(\text{vap.})} = \mu^0 + RT \ln p^0 \quad \text{and} \quad \mu_{(\text{vap.})} = \mu^0 + RT \ln p, \quad (54)$$

assuming them to behave ideally (see p. 304). To allow for departure from ideal behavior the partial vapor pressures should be replaced by the corresponding fugacities, but this refinement can be neglected. Since the liquids and vapor are in equilibrium, the values of $\mu_{0(\text{vap.})}$ and $\mu_{(\text{vap.})}$ given by (54) are also the chemical potentials of the solvent in pure solvent and solution, respectively, and so (53) can be written

$$RT \ln \frac{p^0}{p} = \int_{P_0}^P \bar{V}_1 dP. \quad (55)$$

This equation is thermodynamically exact, except for the fact that fugacities should be employed on the left-hand side. To integrate (55) it may be assumed, for simplicity, that \bar{V}_1 is independent of pressure; that is to say, the solution is taken to be incompressible. It then follows that

$$RT \ln p^0/p = \bar{V}_1(P - P_0). \quad (56)$$

Since $P - P_0$ is, by definition, equal to the osmotic pressure Π , it is seen that

$$\Pi \bar{V}_1 = RT \ln p^0/p, \quad (57)$$

which should apply if the solvent vapor behaves ideally. The accuracy of (57) may be tested by comparing the measured osmotic pressures of various solutions with the values calculated from the vapor pressures. The available data are limited, but some observed and calculated osmotic pressures of sucrose solutions at 30° c. are recorded in Table 112. The

TABLE 112. OBSERVED AND CALCULATED OSMOTIC PRESSURES OF SUCROSE SOLUTIONS AT 30° C.

Conc. Moles per 1000 g.	Observed	Calculated	Conc. Moles per 1000 g.	Observed	Calculated
0.1	2.47 atm.	2.47 atm.	3.0	95.16 atm.	96.2 atm.
1.0	27.22	27.0	4.0	138.96	138.5
2.0	58.37	58.5	5.0	187.3	183.0

close agreement of the observed and calculated osmotic pressures shows that equation (57) is in harmony with experiment over a considerable range of concentration.

The calculation of the partial molar volume \bar{V}_1 of the solvent may be carried out as follows (see also p. 239). If a solution contains n_1 moles of solvent of molecular weight M_1 and 1 mole of solute of molecular weight M_2 in V liters,

then its density ρ is given by the expression

$$\rho = (n_1 M_1 + M_2)/1000V, \quad (58)$$

$$\therefore n_1 = (1000\rho V - M_2)/M_1. \quad (59)$$

On differentiation this gives

$$dn_1 = 1000(\rho dV + Vd\rho)/M_1, \quad (60)$$

$$\therefore \frac{dn_1}{dV} = 1000(\rho + Vd\rho/dV)/M_1. \quad (61)$$

If c is the concentration of the solution in moles per liter, then $c = 1/V$, and so

$$\frac{dn_1}{dV} = 1000(\rho - cd\rho/dc)/M_1, \quad (62)$$

$$\therefore \bar{V}_1 = \frac{dV}{dn_1} = M_1/1000(\rho - cd\rho/dc). \quad (63)$$

To evaluate \bar{V}_1 it is necessary, therefore, to determine the density of the solution at various concentrations, so that the value of $d\rho/dc$ may be obtained from the slope of the curve in which ρ is plotted against c ; the point for measuring the slope is that for the concentration of the solution under consideration.

Since the chemical potentials in (54), which were employed in (55) for the deduction of (57), refer to 1 mole of the solvent as *vapor*, it follows that the partial molar volume \bar{V}_1 must apply to the same quantity of solvent; it is, therefore, the change in volume when the quantity of liquid which gives 1 mole of solvent vapor is added to a large quantity of solution. This emphasis on the mole of solvent as referring to the molecular weight in the vapor is of great importance, as will be seen shortly.

Osmotic Pressure and Concentration.—If the assumption is made that Raoult's law is applicable, that is, by (5), $p/p^0 = x_1 = 1 - x_2$, where x_2 is the mole fraction of the solute, then (57) becomes

$$\Pi \bar{V}_1 = -RT \ln(1 - x_2). \quad (64)$$

If, further, the solution is dilute then $\ln(1 - x_2) \approx -x_2$, and hence

$$\Pi \bar{V}_1 = RTx_2; \quad (65)$$

at the same time x_2 may be replaced by n_2/n_1 , and so

$$\Pi \bar{V}_1 = n_2 RT/n_1. \quad (66)$$

If the density of the solution is a linear function of the molar concentration, which is generally true in dilute solutions, then the quantity $\rho - cd\rho/dc$ in (63) is equal to the density of the solvent. It is evident that under these conditions \bar{V}_1 becomes identical with the molar volume of the pure solvent, and so $n_1 \bar{V}_1$ may be replaced by V' , the volume of solvent associated with n_2 moles of solute in the solution. Equation (66) then takes the form

$$\Pi V' = n_2 RT. \quad (67)$$

This is sometimes known as Morse's equation, as it was proposed empirically by H. N. Morse (1905) as a modification of the van't Hoff equation which was in better agreement with actual experiment (see Table 113 below). By a still further simplification, applicable to an extremely dilute solution, the volume V' of solvent may be replaced by the volume V of the solution, then

$$\Pi V = n_2 RT \quad \text{or} \quad \Pi = cRT, \quad (68)$$

where c , equal to n_2/V , is the concentration of the solution in moles per liter; this is, of course, identical with the van't Hoff equation (50). From the deduction given above it is clear that (64) can apply only to a solution obeying Raoult's law, and further (67) and (68) will hold only when the solutions are very dilute; it is of interest, therefore, to see how far they represent the behavior of actual solutions. The osmotic pressure measurements in the previous table are repeated in Table 113

TABLE 113. OSMOTIC PRESSURES OF SUCROSE SOLUTIONS AT 30° C.

Concentration Moles per 1000 g.	Observed Osmotic Pressure	Ideal	Calculated Morse Equation	van't Hoff Equation
0.1	2.47 atm.	2.44 atm.	2.47 atm.	2.40 atm.
1.0	27.22	24.40	24.72	20.4
2.0	58.37	48.32	49.43	35.1
3.0	95.16	71.85	74.15	45.5
4.0	138.96	94.80	98.86	55.7
5.0	187.3	117.7	123.58	64.5

together with the values calculated for an ideal solution [equation (64)], from the Morse equation (67), and from van't Hoff's equation (68). It will be observed that except in very dilute solutions the departure from the ideal of Raoult's law is quite considerable; the Morse equation apparently introduces, accidentally, a slight compensation, but the osmotic pressures calculated by van't Hoff's relationship differ very greatly from the observed values at appreciable concentrations. This is to be anticipated for, as shown above, and as indicated by van't Hoff, the equation can only be expected to apply to very dilute solutions.

Osmotic Pressure and Other Properties.—For a dilute solution p^0/p is very close to unity, so that by McLaurin's theorem

$$\ln \frac{p^0}{p} \approx \frac{p^0 - p}{p} \approx \frac{p^0 - p}{p^0}. \quad (69)$$

Substitution in (57) then gives, for a dilute solution,

$$\frac{\Pi \bar{V}_1}{RT} = \frac{p^0 - p}{p^0}. \quad (70)$$

As seen above, for such a solution, \bar{V}_1 may be replaced by the molar volume of the solvent, and since this is a constant, and so also are R and T , it

follows that the relative lowering of the vapor pressure is proportional to the osmotic pressure of a dilute solution in a given solvent at constant temperature.

It can be shown (pp. 635, 644) that

$$\ln \frac{p^0}{p} = \frac{L}{R} \cdot \frac{\Delta T}{T_0^2} \quad (71)$$

holds both for the elevation of the boiling point or the depression of the freezing point (ΔT), where L is the molar heat of vaporization or of fusion, respectively, and T_0 is either the boiling or freezing point of the solvent. It follows from (57), therefore, that

$$\Pi = \frac{LT}{\bar{V}_1 T_0^2} \Delta T, \quad (72)$$

where T is the temperature at which the osmotic pressure is measured. As before, \bar{V}_1 may be assumed constant for dilute solutions, and since T_0 and L may also be taken as constant, it is seen that the lowering of the freezing point and the elevation of the boiling point in a given solvent are proportional to the osmotic pressure of the solution. As a result of this connection between osmotic pressure, on the one hand, and lowering of vapor pressure, elevation of boiling point and depression of the freezing point, on the other hand, the latter properties of a solution have been called "osmotic properties." This term is not altogether satisfactory as it lays too great an emphasis on osmotic pressure; the general attitude in recent years has been to regard the free energy or activity of the solvent molecules, as measured by their vapor pressure, as fundamental, and to consider the various properties mentioned above as manifestations of the lower value in the solution than in the solvent; hence the noncommittal expression "colligative properties" is to be preferred (p. 625). It will be understood, of course, it is only a first approximation to state that these properties depend on the number of solute molecules and not on their nature.

Determination of Molecular Weights.—Since dilute solutions obey van't Hoff's equation, at least approximately, measurement of osmotic pressure should provide another method for the determination of molecular weights in solution. If a given solution contains w_2 g. solute of molecular weight M_2 in a volume V liters of solution, then the concentration c is w_2/M_2V moles per liter, and hence van't Hoff's equation becomes

$$\Pi = w_2 RT / M_2 V \quad \text{or} \quad M_2 = w_2 RT / \Pi V \quad (73)$$

and so M_2 should be determinable. Owing to the difficulty of measuring osmotic pressures the method has only found application in a limited field, namely, for substances of very high molecular weight, e.g., proteins, polymeric compounds, such as natural and synthetic rubbers, and carbohydrate derivatives. For substances of this type, however, the method

has proved of great value, especially as the conditions permit a simplification of technique. Not only are the osmotic pressures of such solutions low, so that a strong membrane is unnecessary, but the molecules are so large that they do not pass through membranes which are permeable to substances such as sugar. Consequently, membranes of collodion or of cellophane have proved to be sufficiently semipermeable for osmotic pressure measurements with such substances in aqueous or nonaqueous solutions. The method in which the liquid-vapor surface acts as a semipermeable membrane (p. 660) has also been applied to determine molecular weights of complex molecules in nonaqueous media. Provided the experimental material is neutral, e.g., a carbohydrate, the presence in solution of substances of low molecular weight generally has no influence, since they pass freely through the membrane by diffusion, and as they distribute themselves equally on both sides they do not affect the osmotic pressure. For proteins, and related compounds, however, unless they are at their isoelectric point, the membrane affects the distribution of an electrolyte and suitable conditions must be employed, as explained later (cf. p. 1261).

If (73) were always applicable, Π/w_2 , known as the "specific" or "reduced" osmotic pressure, should be independent of w_2 , for constant volume of solution and temperature. For high polymeric compounds, in particular, the value is found to increase with increasing concentration; this is not unexpected, since (73) should hold only for extremely dilute solutions. By suitable extrapolation of the plot of Π/w_2 against w_2 , the value of the former at infinite dilution may be obtained, and insertion of the resulting value in (73) gives the required molecular weight.²⁰

Review of Molecular Weight Determination.—The colligative properties of solutions, especially lowering of the freezing point and elevation of boiling point, have been frequently used for determining molecular weights in solution, but there has been much uncertainty as to the particular molecules to which the results refer. It is, therefore, worth while to look into the matter more closely. Since the van't Hoff equation is at least approximately applicable to dilute solutions, it is possible, by reversing the arguments on p. 670, to deduce the relationship

$$\Pi \bar{V}_1 = n_2 RT / n_1, \quad (74)$$

where \bar{V}_1 is the partial molar volume of the solvent, referred to the molecular weight as *vapor* (see p. 670). In this equation, therefore, n_1 must be calculated on the basis of the molecular weight of the solvent in the vapor phase in equilibrium with the solution. By continuing the deduction of p. 670 backward, Raoult's law is obtained, and so it follows that in the equation

$$\frac{p^0 - p}{p^0} = \frac{n_2}{n_1 + n_2} = \frac{w_2/M_2}{w_1/M_1 + w_2/M_2}, \quad (75)$$

M_1 refers to the molecular weight of the solvent as vapor. It can be shown that from the standpoint of Raoult's law there is no fundamental

difference between solute and solvent (p. 710), and so it is probable that M_2 refers equally to the *vapor of the solute* * in equilibrium with the solution.

The abnormal results obtained by F. M. Raoult and A. Recoura (1890) for the lowering of the vapor pressures of various solutes in acetic acid solution near the boiling point were shown by J. H. van't Hoff (1890) to be accounted for by the association of acetic acid *vapor*. The average molecular weight of the solvent calculated from Raoult's law, 1.6×60 , is in excellent agreement with that obtained from the density of the vapor at the boiling point. Although liquid water is known to be associated, the lowering of the vapor pressure is not abnormal; this is because water differs from acetic acid in giving a vapor consisting of simple molecules.

In the cryoscopic and ebullioscopic methods the values obtained also refer to the molecular weight of the solute as vapor, since the equations employed can be deduced, as previously shown, from Raoult's law. It may be proved, however, that the results are independent of the molecular weight of the solvent. On p. 635, for example, the values of L , the molar heat of vaporization, and of M_1 , the molecular weight of the solvent, both refer to the vapor; hence L/M_1 , which appears in (23), is the latent heat (l_s) per gram of solvent irrespective of its molecular complexity in liquid or vapor states. The same argument may be used in connection with the depression of the freezing point.²¹

The Ideal Solution.—If van't Hoff's equation is used to define an ideal solution, then departure from ideal behavior must be interpreted with respect to such factors as are likely to influence the osmotic pressure. Van't Hoff himself was content to write $\Pi V = iRT$, where i , known as the **van't Hoff factor**, was introduced to cover all types of deviation from ideal behavior without any question as to their origin. It is, however, necessary to consider the factors likely to cause these deviations. If the solute bombardment theory of osmotic pressure is adopted, then departure from ideal behavior may be attributed to the volume occupied by the solute molecules and to their mutual attraction, as in the van der Waals treatment of gases. A number of attempts have been made in this direction; for example, O. Sackur (1900), A. W. Porter (1917), and others, have shown that in certain instances the variation of osmotic pressure with dilution over a limited range can be represented by

$$\Pi(V - b) = RT, \quad (76)$$

where b is supposed to correct for the volume of the solute molecules, although it varies appreciably with concentration and temperature. Even if b were constant and the agreement of (76) with experiment better than it is, the significance attached to b would still depend on the acceptance of the solute bombardment hypothesis. Actually the advantage of (76) over van't Hoff's simple equation is merely the result of introducing an

* J. M. Peterson and W. H. Rodebush (1928) have given a more rigid proof that in the cryoscopic method it is the molecular weight of the solute as *vapor* which is obtained.

additional term with an empirical constant, as may be seen by writing $1/c$ for V , when (76) becomes

$$\Pi = \frac{RTc}{1 - bc} \approx RTc(1 + bc). \quad (77)$$

On the whole it would seem preferable to discard osmotic pressure as the property in terms of which to define an ideal solution, and to adopt another which is simpler to understand. For this and other reasons, most physical chemists now employ Raoult's equation, in the place of van't Hoff's equation, as the basis of ideal behavior. Not only is the vapor pressure a property more readily comprehended than osmotic pressure, but as will be seen shortly it is much easier to understand the factors likely to influence the former. Further, it is possible to deduce the Raoult equation by means of simple theoretical considerations which appear to be free from serious objection. At infinite dilution the laws of van't Hoff and of Raoult are identical; the equation $\Pi = RTc$ can be derived for a very dilute solution if Raoult's law is assumed (p. 671), and it is equally possible, as van't Hoff showed, to reverse the procedure. At appreciable concentrations, however, the two laws give quite different results (cf. Table 113), and a choice between them is important. In agreement with van't Hoff's expectation (p. 664) there are no solutions obeying his ideal equation, except at very high dilutions, but there are some instances of mixtures for which Raoult's law is applicable over a very large range of concentration (cf. p. 711); this law can, therefore, be used as a basis for the study of concentrated as well as dilute solutions. An **ideal solution** may consequently be defined as *one which obeys Raoult's law over the whole range of concentration and at all temperatures*. Such solutions are only formed from constituents which mix in the liquid state without heat change* (cf. p. 626) and without volume change; in other words, the heat of dilution must be zero and the partial molar volumes of the constituents are constant at all concentrations.

In the treatment of the present chapter, it has been postulated that Raoult's law is applicable to the *solvent*. It is very frequently true that Raoult's law holds, to a good approximation, in dilute solutions, although deviations occur at higher concentrations. These solutions are clearly not ideal, in view of the definition given above; nevertheless, the various equations for rise of boiling point, lowering of freezing point and osmotic pressure, which have been derived above, will apply to them, in the same region that Raoult's law applies to the solvent. Such solutions are sometimes called **ideal dilute solutions** (cf. p. 685).

Theoretical Basis of Raoult's Equation.—Provided its vapor behaves as an ideal gas, that is, the pressure is exactly proportional to the number of molecules per unit volume, the vapor pressure of a given liquid will be determined by the product of the "escaping tendency" of the molecules and the number available. If the addition of a solute to a given solvent

* For a solid solute allowance must be made for the heat of fusion.

leaves quite unchanged the forces existing between the molecules of the latter, then the tendency for each solvent molecule to leave the liquid and enter the vapor will not be affected in any way. The solution consequently behaves, from the standpoint of vapor pressure, merely as if some of the solvent molecules were replaced by those of solute, the escaping tendency of the former remaining the same as in the pure solvent. It appears, therefore, that the vapor pressure p of the solvent above such a solution will be determined only by the ratio of the number of solvent molecules n_1 to the total number of molecules in the solution $n_1 + n_2$, since this determines the proportion of the liquid consisting of solvent molecules. It is therefore possible to write

$$p = k \frac{n_1}{n_1 + n_2} = kx_1, \quad (78)$$

where k is a proportionality constant and x_1 is the mole fraction of solvent. For the pure solvent x_1 is unity and p becomes p^0 ; it follows then that k is equal to p^0 , and hence

$$p = p^0 x_1, \quad (79)$$

which is identical with (5) deduced from Raoult's relationship.*

It must be emphasized that this deduction is based on the assumption that the presence of solute does not affect the environmental forces of the solvent molecules; it can be shown that this will only be the case, when there is no heat or volume change when the liquid components of the solution are mixed.† One obvious result of these conditions is, as already shown in another connection (p. 626), that Raoult's law can hold only for solutions having zero heat of dilution. The total energy of an ideal gas is independent of its volume (p. 191) and two ideal gases which do not react chemically can be mixed without any thermal or volume change at constant pressure. An ideal solution, as defined above, is analogous since the total energy is independent of the concentration, so that there is no heat change on dilution, and the volume is the sum of the separate volumes of the liquid constituents.

Deviations from Ideal Behavior.—The familiar equations for the elevation of boiling point and depression of freezing point have been derived for dilute solutions on the assumption that Raoult's law applies to them; hence deviations from these equations, as well as deviations from the vapor pressure equation, may be regarded equally as indicating departure from ideal behavior. It will be convenient, therefore, to discuss these deviations together; before considering the results it is of interest to inquire into the factors which may determine such deviations.

There are two main reasons why a given solution may not behave ideally; first, the two constituents may be fundamentally incapable of forming an ideal system since the molecules of each affect the intermolecu-

* A more exact treatment suggests that the molecules of solvent and solute should have the same size if Raoult's law is to hold (Guggenheim, 1937).

† When Raoult's law is obeyed there is no heat or volume change, but the reverse of this is not necessarily, although it is generally, true.

lar forces, and hence the escaping tendency, of the other; and second, on account of compound formation between solute and solvent, or as a result of association of either to form complex molecules, the molecular species actually present are not those assumed in making up the solution. These two factors will be examined in turn.

Internal Pressure and Polarity.—It can be deduced thermodynamically that only those liquids which have identical values of $(\partial P/\partial T)_V$ can obey Raoult's law at all temperatures and pressures and hence form ideal mixtures (J. H. Hildebrand, 1921). It was seen on p. 479 that $T(\partial P/\partial T)_V$ is an approximate measure of the internal pressure of a liquid, and so it follows that only substances with equal internal pressures are able to yield ideal solutions. The internal pressure of a liquid is a measure of the attractive force between the molecules; hence when two liquids of equal internal pressure form a liquid solution, it is reasonable to suppose that the environmental forces of each will remain almost unchanged. Such a system might, on general grounds, be expected to obey Raoult's law. If the internal pressures of solute and solvent are different, however, deviations from ideal behavior must result, and it can be shown theoretically that these deviations should be of a positive nature; that is to say, the vapor pressure of each constituent should be greater than for an ideal solution. Provided other factors do not interfere, the extent of the deviations is believed to be very roughly proportional to the difference in the internal pressures of the two components of the solution. If one of the molecular species is polar—the molecule as a whole may have a dipole moment, or it may have no resultant moment but contain polar bonds separated by an appreciable distance, e.g., *p*-dichlorobenzene but not carbon tetrachloride—whereas the other is nonpolar, there will be a change in the intermolecular forces on mixing which is superimposed on that due to differences of internal pressure. The effects produced in this manner act in the same direction, causing positive deviations, but they are of a much smaller order of magnitude. An extreme case of deviations resulting from polarity arises when the solute is an ionizable substance; electrolytic solutions, other than the most dilute, show appreciable departure from ideal behavior even after making allowance for the number of particles produced by dissociation of the solute. Such solutions have, of course, appreciable heats of dilution; this subject will be considered more fully in Chapter XII. A solution containing a long-chain solute dissolved in a simple solvent will also show positive deviations from Raoult's law; the environment of the long molecule is completely changed and it has consequently a high escaping tendency, as is evident from its low solubility, e.g., long-chain fatty acids and alcohols in water.

Compound Formation and Association.—Deviations from Raoult's law occur when the solute and solvent combine to form definite compounds, e.g., hydrates in aqueous solution; the number of free, i.e., uncombined, molecules of both solvent and solute will be less than would have been the case had no compound formation occurred. If the escaping tendency of each molecular species remains unchanged, the vapor pressure

of the solvent in the solution will be smaller than the theoretical value; in other words, compound formation between solute and solvent results in negative deviations from Raoult's law.* It is of interest to note that it is possible for the deviations accompanying compound formation to be almost exactly compensated by the positive deviations resulting from differences of internal pressure; the system then behaves as if it were ideal, although it is actually not so. Such solutions, of which the system ethyl acetate-water is an example (J. Kendall, 1925), have been called **pseudo-ideal**; the depression of the freezing point of water by ethyl acetate follows the ideal curve right up to saturation. It must be mentioned that although this system is pseudo-ideal in the vicinity of 0° c., it may not be so at other temperatures.

Association of either solvent or solute leads to the solution having a vapor pressure higher than would otherwise be expected. If the solute is associated, for example, the mole fraction of the solvent is larger than that calculated on the basis of simple molecules, so that its vapor pressure is greater than required by Raoult's law; the system thus exhibits positive deviations. For an associated solvent the position is somewhat more complicated. The vapor pressure may be regarded as being due mainly to simple molecules present in equilibrium with the associated form; in the solution, where the concentration is less, the degree of association will be less than in the pure solvent, in accordance with the law of equilibrium (cf. p. 843). The ratio of the vapor pressure of the solvent over the solution to that of pure solvent, i.e., p/p^0 , will be greater than the apparent mole fraction of solvent, to which it should be equal by Raoult's law. Association of the solvent should thus also be accompanied by positive deviations from ideal behavior.

It may be pointed out in conclusion that all the deviations mentioned, whatever their cause, only become appreciable as the concentration of solute increases. At infinite dilution all solutions should behave ideally, just as all gases do at infinitesimally small pressures. For very dilute solutions the deviations of the solvent from Raoult's law are, therefore, negligible but they increase with increasing concentration.²²

Deviations and Molecular Weights.—For present purposes deviations from ideal behavior may be tested by determinations of molecular weights of nonelectrolytes, or of feebly ionized compounds, by any of the colligative properties discussed in this chapter. It has been seen (p. 674) that if the vapor of the solvent is associated, the results are abnormal but can be readily accounted for by making allowance for the association. With solvents such as water or alcohols, which are only associated as liquid but not as vapor, or with nonassociated solvents, e.g., benzene, carbon tetrachloride, carbon disulfide or ether, abnormalities of this type do not occur. If association of the solvent is not involved, then it can be seen that positive deviations from Raoult's law—that is to say, p is greater than the theoretical value—will result in the calculated mole frac-

* The compound is, in general, much less volatile than its component parts.

tion of solute being too low, and consequently the apparent molecular weight will be too high. Apart from compound formation, all departures from ideal behavior lead to positive deviations, and so the molecular weights obtained from vapor pressure measurements will generally be higher than expected. In the cryoscopic and ebullioscopic methods, as generally employed, the results should be independent of association of the solvent, as shown above (p. 674), and this has been amply verified by experiments with water and alcohols as solvents. All other factors causing positive deviations from Raoult's law will here also give molecular weights which are too high. In those cases in which definite association of the solute occurs the determined molecular weights will, of course, be higher than normal, and it is not always a simple matter to decide whether the results obtained are due to real association or to other causes.

Carboxylic acids are well known to give high molecular weights in nonhydroxylic solvents, e.g., benzene, nitrobenzene, etc. The following results (Table 114) were obtained for solutions of acetic acid in benzene

TABLE 114. APPARENT MOLECULAR WEIGHT OF ACETIC ACID IN BENZENE SOLUTIONS

Acetic Acid per 1000 g. Benzene	ΔT_f	M	Acetic Acid per 1000 g. Benzene	ΔT_f	M
0.201 g.	0.0156°	65.8	14.25 g.	0.608°	119.6
0.399	0.0277°	73.5	30.57	1.254°	124.5
0.895	0.0539°	84.6	61.44	2.410°	130.2
2.894	0.1472°	100.5	97.56	3.644°	133.3
5.802	0.253°	117.0	148.86	5.202°	145.6

by the cryoscopic method (J. M. Peterson and W. H. Rodebush, 1928; C. R. Bury and H. O. Jenkins, 1934). The molecular weights have been calculated by (25) on the assumption that K_f remains constant at 5.10, although it actually decreases with increasing depression (p. 645). The results, plotted in Fig. 155, of the ratio of the observed molecular weight

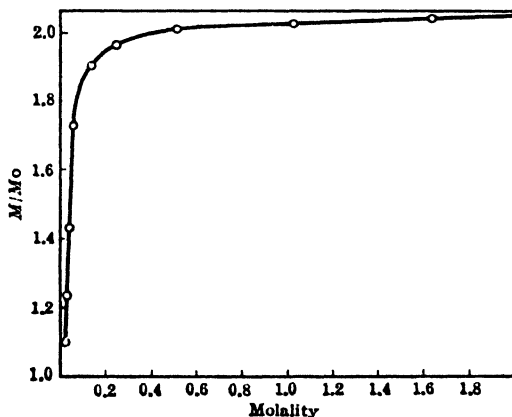
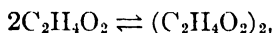


FIG. 155. Molecular weight of acetic acid in benzene solutions

M to the theoretical value M_0 (60.0), show that the vapor * of acetic acid in equilibrium with a very dilute solution in benzene has a normal molecular weight, which increases rapidly until at a concentration of about 0.4 M , it is double the value for simple molecules. This behavior undoubtedly means that acetic acid forms associated molecules; it is in fact to be expected if there is an equilibrium between single and double molecules, thus



for the requirements of the law of equilibrium would necessitate a rapid increase in the proportion of single molecules as the dilution is increased. Attempts have been made in analogous cases to calculate from the apparent molecular weight the proportion of the molecules existing in the two forms. If α is the fraction of the total solute present as double molecules of molecular weight $2M_0$, and $1 - \alpha$ is left in the form of simple molecules, molecular weight M_0 , then the mean, i.e., observed, molecular weight M should be given by

$$M \left(1 - \frac{\alpha}{2} \right) = 2M_0 \left(\frac{\alpha}{2} \right) + M_0(1 - \alpha), \quad (80)$$

$$\therefore \alpha = 2(M - M_0)/M. \quad (81)$$

It is doubtful, however, if the figures obtained in this manner give anything more than an approximate indication of the extent of double molecule formation. Apart from the fact that they apply theoretically to the vapor of the solute, the results do not comply with the law of equilibrium. The chief reason for this is that the calculations of molecular weight assume the solutions to be ideal, whereas actually they cannot be so, on account of differences of internal pressure, polarity, etc., of solute and solvent. The departure from ideal behavior is also shown by the molecular weights, as in Table III, being more than twice the value for simple molecules in the more concentrated solutions, although it is unlikely that any species more complex than double molecules will be present. Even after making allowance for the decrease in K_f at the lower temperatures, the discrepancy still remains. Other carboxylic acids give results similar to those described above; benzoic acid has a double molecular weight in solutions of quite low concentration, e.g., 1.6 g. per 1000 g, benzene, but it is probable that in sufficiently dilute solutions the value would decrease as with acetic acid. It is evident, especially in view of the low vapor pressure of benzoic acid in equilibrium with benzene solutions, that the double molecules must be very stable.

With hydroxylic solvents, e.g., water and alcohols, carboxylic acids give molecular weights which are very little different from those expected for simple molecules. Such solvents are said to be **dissociating solvents**,

* The view commonly held that the results refer to the actual molecules in the solution is probably incorrect (p. 674). It is obvious, however, that if the values show associated molecules to be present in the vapor, they will almost certainly exist in the solution in equilibrium with it, probably to a greater, but actually unknown, extent.

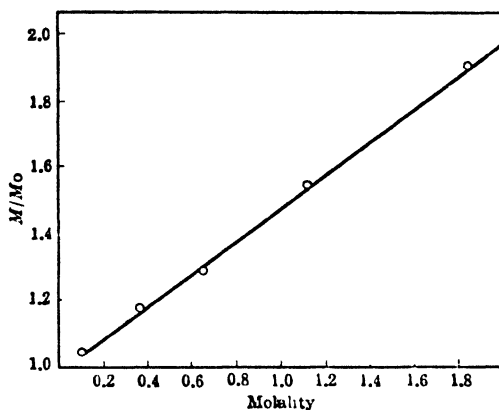
for it appears that they alter the position of equilibrium between single and double molecules in favor of the former. This property may be connected with the relatively high dielectric constants of the solvents and also with the fact that there will be some tendency for the single molecules of the carboxylic acid to combine with solvent molecules as an alternative to association; since there is a large concentration of the latter in the solution, the effect may be considerable. In contrast to the behavior of hydroxylic solvents, some writers have referred to benzene, other hydrocarbons and similar substances as "associating solvents." This term is not altogether justifiable, for the solutes are already associated, both in liquid and vapor states, and the solvent does not contribute in any way to the extent of the association. In fact its diluting effect may result in dissociation into simple molecules, as seen in Fig. 155 for acetic acid.

An abnormality of a different type from that considered above is shown by hydroxylic substances other than acids, e.g., alcohols and phenols, in nonhydroxylic solvents; this may be seen from the results in

TABLE 115. APPARENT MOLECULAR WEIGHTS OF *p*-CRESOL IN BENZENE SOLUTIONS

<i>p</i> -Cresol per 1000 g. Benzene	ΔT_f	M	<i>p</i> -Cresol per 1000 g. Benzene	ΔT_f	M
9.30 g.	0.420°	113.3	69.00 g.	2.539°	139.2
25.27	1.110°	117.6	123.4	3.773°	167.6
41.76	1.663°	128.6	200.0	5.002°	204.7

Table 115 (C. R. Bury and H. O. Jenkins, 1934) for *p*-cresol in benzene, the ratio of the experimental (M) to the theoretical (M_0) molecular weight, i.e., 108.1, being plotted in Fig. 156. The regular increase in the observed

FIG. 156. Molecular weight of *p*-cresol in benzene solutions

molecular weight with increasing concentration suggests that simple association, of the kind found in carboxylic acids, does not occur. Part of the variation is no doubt to be ascribed to the differences in internal pressure and polarity of the solvent and solute, but this will not account for the

whole. It is most probable that *p*-cresol and other hydroxylic substances are associated in solution, and are in equilibrium with associated vapors, but the association is of a general type and does not necessarily lead to the formation of definite polymers (cf. p. 115). The densities of these compounds as vapors show that there is relatively little association at the boiling point; it may nevertheless be appreciable at lower temperatures. In hydroxylic solvents, e.g., water and acetic acid, alcohols and phenols give normal molecular weights; the differences in internal pressure are, of course, less and association diminishes, possibly as a result of complex formation between solvent and simple molecules of solute.

In between cases such as alcohols and phenols, on the one hand, and those exhibiting almost normal molecular weights, on the other hand, there are instances of behavior which are not easy to interpret. An example of this type is nitrobenzene in benzene, for which the data (L. G. Davy and N. V. Sidgwick, 1933) are given in Table 116, and plotted in

TABLE 116. APPARENT MOLECULAR WEIGHT OF NITROBENZENE IN BENZENE SOLUTIONS

Conc. Moles/Liter	ΔT_f	M	Conc. Moles/Liter	ΔT_f	M
0.1200	0.689°	126.7	0.7357	3.945°	143.7
0.2862	1.632°	129.7	1.178	6.225°	152.7
0.4945	2.727°	136.9	1.526	8.015°	159.5

Fig. 157; the theoretical molecular weight is 123. The possibility that the high molecular weights might be due to association has been considered, but it is more probable that the deviations from the normal value are to be ascribed mainly to marked departure of the system from ideal behavior because of the differences of internal pressure, etc., of the two components. There may be some dipole association, but it is probably small (cf. p. 550) and it is impossible to estimate the extent from the measurements recorded above. Unless the internal pressures of solute and solvent are very different, the error in molecular weight determina-

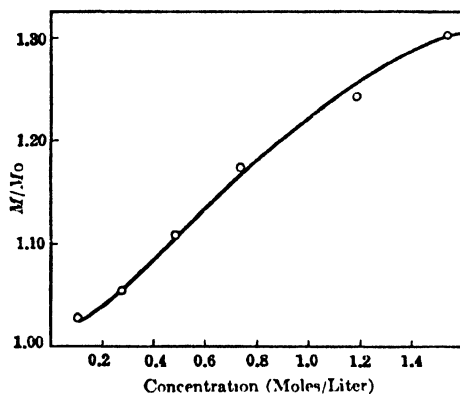


FIG. 157. Apparent molecular weight of nitrobenzene in benzene solutions

tions due to this factor is generally not large. Consider, for example, a solution containing 1 mole of solute dissolved in 1000 g. of solvent of molecular weight 100; this would be a relatively concentrated solution for molecular weight purposes, but the mole fraction of solute is only 0.09, so that deviations from ideal behavior will not be marked. Where the intermolecular forces of solvent and solute differ considerably, however, as is the case with nitrobenzene and cyclohexane, marked abnormalities are observed. For example, the molecular weight of nitrobenzene, as determined by the cryoscopic method with a molar solution in cyclohexane, is approximately twice the normal value; this is not due to association, but to departure from ideal behavior resulting from a large difference in the internal pressures. Confirmation of this view is obtained from the observation that at a temperature slightly below the freezing point of cyclohexane the system separates into two metastable liquid layers; partial miscibility of the two liquids is a clear sign of considerable positive deviation from Raoult's law (p. 711). Somewhat similar results have been obtained for phosphorus in carbon disulfide by the ebullioscopic method; the molecular weight in relatively concentrated solutions is greatly in excess of the expected value for P_4 molecules. The system is, however, far from ideal as is shown by the formation of two (metastable) liquid layers at -6.5°C . (J. H. Hildebrand and T. F. Buerher, 1920). These facts should be a warning against the rash interpretation of the apparent molecular weights obtained from measurements in solution.

Another possible source of error in molecular weight determinations, especially in aqueous solutions, is the formation of compounds, e.g., hydrates, between solute and solvent. At one time it was common for deviations from ideal behavior to be ascribed to this factor, and many calculations have been made of the so-called extent of hydration of substances in aqueous solution. With the gradual realization of the existence of other factors the question of hydration has received less attention, but it is nevertheless a matter of importance which cannot be ignored.²³

THERMODYNAMICS OF NONIDEAL SOLUTIONS

Activity and Activity Coefficients.—It is possible by means of thermodynamic methods to derive equations for nonideal solutions and to express the departure from ideality in a form which is quite independent of the factors responsible. If a liquid and vapor are in equilibrium the chemical potential of a given component will be the same in both phases (p. 477); it follows, therefore, from (185), p. 301, that the chemical potential of any component in a liquid solution may be represented by

$$\mu = \mu^0 + RT \ln p, \quad (82)$$

where p is the partial pressure or, more correctly, the fugacity, of the component in the vapor in equilibrium with the solution. If the latter is

ideal, then the partial pressure, or fugacity, is proportional to the mole fraction (x) of that constituent of the solution; * hence (82) becomes

$$\mu = \mu_x^0 + RT \ln x. \quad (83)$$

It should be noted that μ_x^0 for the given substance is a constant, independent of the composition of the solution, but depending on the temperature and pressure, for the relationship between the vapor pressure and mole fraction of the components depends on both temperature and total pressure.

For a nonideal solution it is necessary to modify (83) by introducing an activity coefficient, f (G. N. Lewis, 1901); thus,

$$\mu = \mu_x^0 + RT \ln xf, \quad (84)$$

where f approaches unity as the system tends toward ideal behavior. The product xf is called the **activity** of the particular component and is represented by the symbol a , viz.,

$$a = xf, \quad (85)$$

so that,

$$\mu = \mu_x^0 + RT \ln a. \quad (86)$$

Since the activity equation is logarithmic in form, it is apparent that the activity of any substance can be expressed only in terms of a *ratio* of the two chemical potentials μ and μ_x^0 . It is the practice, therefore, to choose a reference state, or **standard state**, in which the activity of the given substance is *arbitrarily taken as unity*. The activity of a component in any solution is thus expressed as the ratio of its value to that in the chosen standard state. The particular nature of the latter depends on circumstances, as will be seen below. Whatever the standard state, however, the chemical potential will be equal to the corresponding value of μ^0 , by (86), and this is constant at a definite temperature and external pressure.

It is known that as a solution approaches infinite dilution the behavior of the *solvent* becomes more nearly ideal; that is, the fugacity (or vapor pressure) is more closely proportional to the mole fraction (cf. p. 712). In other words, the activity coefficient f_1 of the solvent tends to unity as x_2 , the mole fraction of solute, approaches zero, and x_1 , the mole fraction of solvent, approaches unity. Consequently, $f_1 x_1$, which is equal to the activity a_1 , approaches unity as the solution becomes less and less concentrated and tends toward pure solvent. It is, therefore, the invariable practice to choose the pure liquid solvent as the standard state of unit activity. For a system consisting of two completely miscible liquids, such as are considered in Chapter X, solvent and solute are indistinguishable, and then the standard state for each component of the mixture is chosen as the respective pure liquid.

For a solution of a solid solute in a liquid solvent, the standard state of the latter is as described above, but for the *solute* it is more convenient to

* This will also be true for a solute which obeys Henry's law (see p. 701).

choose an entirely different standard state. In a *dilute solution* the mole fraction of the solute is proportional to its concentration or molality (see p. 822), and so for such a solution it is possible to write three alternative forms of (83), viz.,

$$\mu = \mu_x^0 + RT \ln x, \quad (87a)$$

$$\mu = \mu_c^0 + RT \ln c, \quad (87b)$$

$$\mu = \mu_m^0 + RT \ln m, \quad (87c)$$

where x , c and m are the mole fraction, concentration and molality, respectively, of the solute; μ_x^0 , μ_c^0 and μ_m^0 are constants, at definite temperature and pressure, that are different but related to one another. For solutions of appreciable concentration, which do not behave ideally, it is necessary to include the appropriate activity coefficients, so that

$$\mu = \mu_x^0 + RT \ln x f_x = \mu_x^0 + RT \ln a_x, \quad (88a)$$

$$\mu = \mu_c^0 + RT \ln c f_c = \mu_c^0 + RT \ln a_c, \quad (88b)$$

$$\mu = \mu_m^0 + RT \ln m f_m = \mu_m^0 + RT \ln a_m, \quad (88c)$$

where the a terms are the respective activities. The standard state of the solute is now chosen so that *the activity coefficient is unity at infinite dilution*. The activity is thus defined in three alternative ways, namely that it becomes equal (a) to the mole fraction, (b) to the concentration, or (c) to the molality, of the solute, as infinite dilution is approached. The three corresponding activity coefficients will in general be different, for any particular solution, although in accordance with the defined standard states they must all approach unity as the solution becomes more dilute (see p. 961).

It will be noted that when f_x is unity the activity of the solute is equal to its mole fraction. For a solution which obeys the ideal form of Henry's law (p. 701) the vapor pressure of the solute is proportional to its mole fraction, and hence the latter may be identified with its activity. It follows, therefore, that the deviation of f_x from unity represents the departure of the solute from Henry's law, or, in general, from the behavior of an "ideal dilute solution" (pp. 675, 712). Provided the solution is reasonably dilute, c and m are approximately proportional to x , and so f_c and f_m also indicate the departure from the behavior of an ideal dilute solution. In solutions of appreciable concentration, however, f_c and f_m will not be unity even for a solution to which Henry's law is applicable.

The Osmotic Coefficient.—It is sometimes convenient to represent departure from ideal behavior in another manner (N. Bjerrum, 1909), by writing

$$\mu = \mu_x^0 + gRT \ln x_1, \quad (89)$$

where x_1 is the mole fraction of the solvent; the deviation factor g is called the **osmotic coefficient**. Consideration of the derivation of (57) shows

that g is equal to the ratio of the observed to the ideal osmotic pressure of the solution. By comparing (84) and (89) it is seen that

$$\ln f_1 = (g - 1) \ln x_1, \quad (90)$$

which gives a relationship between the osmotic coefficient and the activity coefficient of the solvent. The latter may now be related to the activity coefficient of the solute by the Gibbs-Duhem equation (159), p. 238, which for a binary solution can take the form $x_1 d\mu_1 + x_2 d\mu_2 = 0$; at constant temperature and pressure this leads to the relationship

$$x_1 d \ln f_1 + x_2 d \ln f_2 = 0. \quad (91)$$

Introducing the value of $d \ln f_1$ given by (90),

$$x_1 d[(g - 1) \ln x_1] + x_2 d \ln f_2 = 0. \quad (92)$$

For a dilute solution, when x_2 is small, $\ln x_1 = \ln(1 - x_2) \approx -x_2$; making this substitution for $\ln x_1$, and setting x_1 equal to unity, which is justifiable for a dilute solution, (92) becomes

$$d[(1 - g)x_2] + x_2 d \ln f_2 = 0.$$

Further, since the solution is dilute, the mole fraction of the solute is proportional to its molality m , so that it is possible to write

$$d[(1 - g)m] + m d \ln f_2 = 0, \quad (93)$$

where g has been replaced by g , sometimes called the "practical" osmotic coefficient. The significance of this quantity may be seen on p. 966, where an expression similar to (93) is used to derive activity coefficients from freezing-point measurements.

Variation of Activity with Temperature.—By combining (113) and (121) of Chapter III, there is obtained

$$F - T \left(\frac{\partial F}{\partial T} \right)_P = H \quad (94)$$

and differentiating with respect to n , the number of moles of any one component, the others being kept constant, it follows that

$$\mu - T \left(\frac{\partial \mu}{\partial T} \right)_{P, n_1, \dots} = \bar{H}, \quad (95)$$

where μ is equal to $\partial G / \partial n$ and \bar{H} , the partial molar heat content of the component, is $\partial H / \partial n$. Dividing both sides of (95) by T^2 it is seen that the left-hand side is now equivalent to $-\partial(\mu/T) / \partial T$, and hence

$$\left(\frac{\partial(\mu/T)}{\partial T} \right)_{P, n_1, \dots} = -\frac{\bar{H}}{T^2}, \quad (96)$$

which, like (94) and (95), is a form of the Gibbs-Helmholtz equation. For an infinitely dilute solution the chemical potentials of solvent ($\mu_{1(0)}$)

and solute ($\mu_{2(0)}$) are given by

$$\mu_{1(0)} = \mu_1^0 + RT \ln x_{1(0)} \quad \text{and} \quad \mu_{2(0)} = \mu_2^0 + RT \ln x_{2(0)} \quad (97)$$

and for a solution of appreciable concentration,

$$\mu_1 = \mu_1^0 + RT \ln x_1 f_1 \quad \text{and} \quad \mu_2 = \mu_2^0 + RT \ln x_2 f_2. \quad (98)$$

Subtracting each pair of corresponding equations, and dividing by T ,

$$\frac{\mu_1}{T} - \frac{\mu_{1(0)}}{T} = R \ln x_1 f_1 / x_{1(0)} \quad \text{and} \quad \frac{\mu_2}{T} - \frac{\mu_{2(0)}}{T} = R \ln x_2 f_2 / x_{2(0)}. \quad (99)$$

Differentiating with respect to T , the concentrations being constant, and utilizing (96) it is seen that, at constant pressure,

$$\frac{\partial \ln f_1}{\partial T} = - \frac{\bar{H}_1 - \bar{H}_{1(0)}}{RT^2} \quad \text{and} \quad \frac{\partial \ln f_2}{\partial T} = - \frac{\bar{H}_2 - \bar{H}_{2(0)}}{RT^2}. \quad (100)$$

Since $\bar{H}_{1(0)}$ and $\bar{H}_{2(0)}$ refer to infinite dilution they are equivalent to H_1^0 and H_2^0 on p. 241, and hence (100) becomes

$$\frac{\partial \ln f_1}{\partial T} = - \frac{\bar{L}_1}{RT^2} \quad \text{and} \quad \frac{\partial \ln f_2}{\partial T} = - \frac{\bar{L}_2}{RT^2}, \quad (101)$$

where \bar{L}_1 and \bar{L}_2 are the relative partial molar heat contents of solvent and solute, respectively.* These may be evaluated from measurements of the heat of dilution (p. 242). Since the concentration is constant in a given solution, equations similar to (101) also represent the variation of activity with temperature.

Determination of Activities: The Vapor Pressure Method.—The simplest method for evaluating the activity of a solvent is by determination of vapor pressure. If the vapor is assumed to behave as an ideal gas, then both (82) and (86) represent the chemical potential; it follows, therefore, that the activity (a_1) of the solvent in a solution is proportional to its partial vapor pressure (p_1), i.e., $a_1 = kp_1$. It has been seen that in the pure liquid the activity of the solvent is unity, and since the vapor pressure is then p_1^0 , it follows that $k = 1/p_1^0$, and hence

$$a_1 = p_1/p_1^0, \quad (102)$$

so that the activity of the solvent can be obtained from its partial vapor pressure and that of the pure liquid.

If the solute is volatile, its activity may be derived in a somewhat similar manner. The relationship $a_2 = kp_2$ is applicable, and the constant k is determined by utilizing the fact that the activity becomes equal to the mole fraction in very dilute solutions. If $a_{2(0)}$ is the activity, $x_{2(0)}$ the mole fraction and $p_{2(0)}$ the partial vapor pressure of the solute in such a

* For an ideal solution f_1 and f_2 are always unity, and hence \bar{L}_1 and \bar{L}_2 must be zero; there can consequently be no heat change accompanying the addition or removal of solvent. The heat of dilution of an ideal solution is, therefore, zero at all concentrations.

solution, then $a_{2(0)} = x_{2(0)} = kp_{2(0)}$. It follows, therefore, that

$$k = x_{2(0)}/p_{2(0)}$$

and hence

$$a_2 = p_2(x_{2(0)}/p_{2(0)}). \quad (103)$$

It is necessary, therefore, to measure the partial vapor pressure of the solute over a very dilute solution, as well as that over the experimental solution. Since the former cannot be obtained with any accuracy, the method is seldom used. It is possible, however, to compare the activities of a volatile solute in two solutions of different concentrations by means of their respective vapor pressures; if the activity of the solute in one solution is known from other sources that in the other can be calculated. This method has been employed for aqueous solutions of hydrochloric acid.

A more convenient procedure is to calculate the activity of the solute, whether it is volatile or not, from the activity of the solvent obtained from vapor pressure data. Since the chemical potential is a partial molar quantity, a form of the Gibbs-Duhem equation (p. 238) for a binary mixture is $n_1 d\mu_1 + n_2 d\mu_2 = 0$, and utilizing (86) it follows that

$$n_1 d \ln a_1 + n_2 d \ln a_2 = 0, \quad (104)$$

where n_1 and n_2 are the numbers of moles of solvent and solute, respectively; hence,

$$d \ln a_2 = - \frac{n_1}{n_2} d \ln a_1. \quad (105)$$

Integrating between the limits represented by solutions of composition n_1 and n_2 , and n'_1 and n'_2 , respectively, then

$$\ln \frac{a_2}{a'_2} = - \int_{n'_1}^{n_1} \frac{n_1}{n_2} d \ln a_1. \quad (106)$$

By plotting n_1/n_2 against $\ln a_1$ the area under the curve between the limits n_1 and n'_1 gives the integral on the right-hand side; hence the ratio of the activities of the solute in the two solutions may be determined.

The graphical integration of (106) is not always accurate and a modified method has been recommended. Since $x_1 + x_2 = 1$ for a binary mixture, it follows that $dx_1 + dx_2 = 0$; hence,

$$x_1 \frac{dx_1}{x_1} + x_2 \frac{dx_2}{x_2} = 0,$$

$$\therefore x_1 d \ln x_1 + x_2 d \ln x_2 = 0 \quad \text{or} \quad n_1 d \ln x_1 + n_2 d \ln x_2 = 0. \quad (107)$$

Subtraction from (104) gives

$$n_1 d \ln a_1/x_1 + n_2 d \ln a_2/x_2 = 0, \quad (108)$$

$$\therefore d \ln \frac{a_2}{x_2} = - \frac{n_1}{n_2} d \ln \frac{a_1}{x_1}. \quad (109)$$

Integrating between the same limits as before,

$$\ln \frac{a_2}{x_2} - \ln \frac{a'_2}{x'_2} = - \int_{n'_1}^{n_1} \frac{n_1}{n_2} d \ln \frac{a_1}{x_1}. \quad (110)$$

If the composition n'_1 and n'_2 represents infinite dilution then a'_2/x'_2 is unity, and hence

$$\ln \frac{a_2}{x_2} = - \int_{\infty}^{n_1} \frac{n_1}{n_2} d \ln \frac{a_1}{x_1}. \quad (111)$$

To determine $\ln a_2/x_2$, or $\log a_2/x_2$, plot $\ln a_1/x_1$, or $\log a_1/x_1$, against n_1/n_2 and measure the area under the curve between the values of $n_1/n_2 = \infty$ and that corresponding to the concentration of the given solution. It is seen, therefore, that if the activity (a_1) of the solvent is known in various solutions that of the solute can be calculated. It will be noted incidentally that since a/x is equal to the activity coefficient (f_x), the equations (108) to (111) really relate these coefficients for solvent and solute. If the latter is an electrolyte the f_x values may be converted into f_i or γ by means of the equations on p. 961.

The Freezing Point Method.—The evaluation of activities from freezing point observations is generally used for electrolytes, and is capable of giving accurate results even in dilute solutions, when the vapor pressure method is not satisfactory on account of the small differences of the pressures from that of the pure solvent. It has been seen that (32) is applicable at the freezing point of a solution, provided the vapor obeys the gas laws; if not, then the pressures should be replaced by the corresponding fugacities. The ratio of the fugacity in a solution to that of the pure solvent may be identified with the activity (a_1) of the solvent in the solution (cf. p. 687); hence (32) may be written

$$\frac{d \ln a_1}{dT} = \frac{L_f}{RT^2}, \quad (112)$$

whether the vapor behaves ideally or not. Assuming L_f to be independent of temperature, (112) may be integrated between T , the freezing point of the solution, when the activity is a_1 , and the freezing point of the pure solvent T_0 , when the activity is unity; the result is

$$\ln a_1 = - \frac{L_f}{R} \left(\frac{1}{T} - \frac{1}{T_0} \right) \quad (113)$$

$$= - \frac{L_f}{R} \cdot \frac{\Delta T_f}{T_0 T}, \quad (114)$$

where ΔT_f , the depression of the freezing point, is substituted for $T_0 - T$. If the solution is reasonably dilute $T_0 T$ is almost equal to T_0^2 , and hence

$$\ln a_1 = - \frac{L_f}{RT_0^2} \Delta T_f. \quad (115)$$

This relationship permits the activity of the solvent in the solution to be determined at the freezing point of the latter.

To derive an equation for the activity of the solute from (115), application is made of (105); again assuming L_f to remain constant,

$$d \ln a_2 = \frac{n_1}{n_2} \cdot \frac{L_f}{RT_f^2} d(\Delta T_f), \quad (116)$$

although this equation holds whether L_f is constant or not, since it has been previously subjected to integration in arriving at (115). If m is the molality of the solution, i.e., m moles of solute in 1000 g. of solvent, then n_1 is equal to $1000/M_1$, where M_1 is the molecular weight of the solvent, and n_2 is equivalent to m ; hence,

$$d \ln a_2 = \frac{1000L_f}{RT_f^2 M_1} \cdot \frac{d(\Delta T_f)}{m}. \quad (117)$$

Since L_f/M_1 is equal to l_f , the latent heat of fusion per g., the first fraction on the right-hand side is identical with the reciprocal of K_f , the freezing-point depression constant (p. 645), and so

$$d \ln a_2 = d(\Delta T_f)/mK_f, \quad (118)$$

for the activity of the solute at the freezing point of the solution. If the variation of ΔT_f with the molality of the solution is determined, it is possible to carry out a graphical integration and so evaluate the activity of the solute; since the method is generally employed in connection with electrolytes, it is discussed more fully in Chapter XII, p. 965.

Activities of solvent and solute can be also determined from measurements of the rise of boiling point, and it is easily possible to derive equations similar to those given above. As ebullioscopic measurements cannot be made with such accuracy as freezing point determinations, the procedure is not often used. Other methods for the evaluation of activities and activity coefficients, with special reference to electrolytes, will be considered in Chapter XII.²⁴

REFERENCES

- Butler, J. A. V. "The Fundamentals of Chemical Thermodynamics." Vol. II.
 Findlay, A. "Osmotic Pressure."
 Guggenheim, E. A. "Modern Thermodynamics."
 Hildebrand, J. H. "Solubility."
 Lewis, G. N., and Randall, M. "Thermodynamics, etc."
 MacDougall, F. H. "Thermodynamics and Chemistry."
 Taylor, H. S. "Treatise on Physical Chemistry," 2nd. ed., Chap. VII (J. C. W. Frazer).
1. Raoult, *Compt. rend.*, **103**, 1125 (1886); *Z. phys. Chem.*, **2**, 353 (1888); *Ann. Chim. Phys.*, **15**, 375 (1888); **20**, 297 (1890); see also van't Hoff, *J. Chem. Soc.*, **81**, 969 (1902), Raoult Memorial Lecture.
 2. Frazer, Lovelace, *et al.*, *J. Am. Chem. Soc.*, **36**, 2439 (1914); **42**, 1793 (1920); **45**, 2930 (1923); Frowein, *Z. phys. Chem.*, **1**, 1 (1887); Brønsted and Richards, *J. Am. Chem. Soc.*, **50**, 3030 (1928); Gibson and Adams, *ibid.*, **55**, 2679 (1933); Menzies, *ibid.*, **33**, 1615 (1911).
 3. Walker, *Z. phys. Chem.*, **2**, 602 (1888); Berkeley and Hartley, *Proc. Roy. Soc.*, **77A**, 156 (1906); *Phil. Trans. Roy. Soc.*, **218A**, 295 (1919); Washburn and Heuse,

- J. Am. Chem. Soc.*, **37**, 309 (1915); Pearce, *et al.*, *J. Phys. Chem.*, **31**, 231 (1927); *J. Am. Chem. Soc.*, **59**, 2689 (1937); Bechtold and Newton, *ibid.*, **62**, 1390 (1940).
4. Cumming, *J. Chem. Soc.*, **95**, 1772 (1909); McBain and Salmon, *J. Am. Chem. Soc.*, **42**, 426 (1920); Hepburn, *J. Chem. Soc.*, 550 (1932); Hixson and White, *Ind. Eng. Chem. (Anal.)*, **10**, 235 (1938).
 5. Barger, *J. Chem. Soc.*, **85**, 286 (1904); Sinclair, *J. Phys. Chem.*, **37**, 395 (1933); Sinclair and Robinson, *J. Am. Chem. Soc.*, **56**, 1830 (1934); Scatchard, *et al.*, **60**, 3061 (1938); Lassetre and Dickinson, *ibid.*, **61**, 54 (1939); Felsing, *et al.*, *ibid.*, **64**, 244 (1942); Roepke, *J. Phys. Chem.*, **46**, 359 (1942).
 6. van't Hoff, *Z. phys. Chem.*, **1**, 481 (1887); *Phil. Mag.*, **26**, 81 (1888).
 7. Beckmann, *Z. phys. Chem.*, **4**, 532 (1889); **8**, 223 (1890); **15**, 656 (1894); **21**, 239 (1896); **64**, 506 (1908); Bigelow, *Am. Chem. J.*, **22**, 280 (1899); Mathews, *et al.*, *Trans. Am. Elec. Soc.*, **19**, 81 (1911); *J. Am. Chem. Soc.*, **47**, 1791 (1925); Menzies, *ibid.*, **43**, 2309, 2314 (1921).
 8. Sakurai, *J. Chem. Soc.*, **61**, 989 (1892); Walker and Lumsden, *ibid.*, **73**, 502 (1898); Landsberger, *Ber.*, **31**, 461 (1898); McCoy, *Am. Chem. J.*, **23**, 353 (1890).
 9. Cottrell, *J. Am. Chem. Soc.*, **41**, 721 (1919); Washburn and Read, *ibid.*, p. 729; Spencer, *ibid.*, **43**, 301 (1921); Pearce and Hicks, *J. Phys. Chem.*, **30**, 1678 (1926); Bancroft and Davis, *ibid.*, **33**, 591 (1929); Saxton and Smith, *J. Am. Chem. Soc.*, **54**, 2626 (1932); Smith, *ibid.*, **61**, 497 (1939); Davis, *J. Chem. Ed.*, **10**, 47 (1933).
 10. Beckmann, *Z. phys. Chem.*, **2**, 638, 715 (1888); **7**, 323 (1891); **15**, 656 (1894); **21**, 239 (1896); Jones, *ibid.*, **11**, 110 (1893); Walker and Robertson, *Proc. Roy. Soc. Edin.*, **24**, 363 (1903); Richards, *J. Am. Chem. Soc.*, **25**, 291 (1903); *Z. phys. Chem.*, **44**, 563 (1903); Adams, *J. Am. Chem. Soc.*, **37**, 481 (1915); Hovorka and Rodebush, *ibid.*, **47**, 1614 (1925); Randall and Scott, *ibid.*, **49**, 647 (1927); Scatchard, *et al.*, *ibid.*, **54**, 2676 (1932); Lange, *Z. phys. Chem.*, **A186**, 291 (1940).
 11. Jouniaux, *Bull. Soc. chim.*, **11**, 722 (1912); Rast, *Ber.*, **55**, 1051 (1922); Le Fèvre and Tideman, *J. Chem. Soc.*, 1729 (1931); Bernstein and Miller, *J. Am. Chem. Soc.*, **62**, 948 (1940); Ricci, *ibid.*, **66**, 658 (1944).
 12. van't Hoff, *Z. physikal. Chem.*, **5**, 322 (1890); van Bijlert, *ibid.*, **8**, 343 (1891); Beckmann, *ibid.*, **17**, 120 (1895); **22**, 612 (1897).
 13. For an account of early developments in the study of osmotic pressure, see Findlay, *op. cit.*
 14. Tinker, *Phil. Mag.*, **32**, 295 (1916); **33**, 428; **34**, 526 (1917); *Proc. Roy. Soc.*, **92A**, 357 (1916); **93A**, 268 (1917); Callendar, *ibid.*, **80A**, 466 (1908); Keggin and Miles, *Nature*, **137**, 577 (1935); Fordham and Tyson, *J. Chem. Soc.*, 483 (1937); McMahon, *et al.*, *Trans. Farad. Soc.*, **36**, 515 (1940).
 15. Berkeley and Hartley, *Phil. Trans. Roy. Soc.*, **206A**, 486 (1906); **209A**, 177 (1909); *Proc. Roy. Soc.*, **82A**, 271 (1909); Frazer, *et al.*, *J. Am. Chem. Soc.*, **38**, 1907 (1916); **43**, 2497 (1921); **45**, 1710 (1923); Murray, *J. Phys. Chem.*, **33**, 896 (1929).
 16. Townsend, *J. Am. Chem. Soc.*, **50**, 2958 (1928); Eichelberger, *ibid.*, **53**, 2025 (1931); Martin and Schultz, *J. Phys. Chem.*, **35**, 638 (1931); Ullmann, *Z. phys. Chem.*, **156**, 419 (1931).
 17. de Vries, *ibid.*, **2**, 415 (1883); Hamburger, *ibid.*, **6**, 319 (1890).
 18. van't Hoff, *loc. cit.*, ref. (6); Ramsay, *Phil. Mag.*, **38**, 206 (1884).
 19. Tinker, *loc. cit.*, ref. (14); Schay, *Z. phys. Chem.*, **106**, 378 (1923); Hulett, *ibid.*, **42**, 353 (1902); Thiel, *ibid.*, **178**, 374 (1937).
 20. Adair, *Proc. Roy. Soc.*, **108A**, 627 (1925); Burk, *J. Biol. Chem.*, **98**, 353 (1932); Ullmann, *Ber.*, **69**, 1426, 1442, 1448 (1936); Oakley, *Trans. Farad. Soc.*, **31**, 136 (1935); **33**, 372 (1937); Carter and Record, *J. Chem. Soc.*, 660, 664 (1939); Hess and Surányi, *Z. phys. Chem.*, **A184**, 321 (1939); Donnan, *J. Chem. Soc.*, 707 (1939), a review; Montonna and Jilk, *J. Phys. Chem.*, **45**, 1374 (1941); Fuoss and Mead, *ibid.*, **47**, 59 (1943); see also, General Discussion, *Trans. Farad. Soc.*, **40**, (1944).

21. Raoult and Recoura, *Z. phys. Chem.*, **5**, 422 (1890); van't Hoff, *ibid.*, p. 425; Rosanoff and Dunphy, *J. Am. Chem. Soc.*, **36**, 1411 (1914); Peterson and Rodebush, *J. Phys. Chem.*, **32**, 709 (1928); Bancroft and Davis, *ibid.*, p. 1; **33**, 361 (1929).
22. Lewis, *J. Am. Chem. Soc.*, **30**, 668 (1908); Washburn, *Z. phys. Chem.*, **74**, 537 (1910); Hildebrand, *J. Am. Chem. Soc.*, **38**, 1452 (1916); *op. cit.*; Guggenheim, *J. Phys. Chem.*, **34**, 1751 (1930); *Proc. Roy. Soc.*, **162A**, 346 (1932); *Trans. Farad. Soc.*, **33**, 151 (1937).
23. Bury and Jenkins, *J. Chem. Soc.*, 688 (1934); Davy and Sidgwick, *ibid.*, 281 (1933); Hildebrand, *op. cit.*, p. 168 (1936).
24. Lewis, *Z. phys. Chem.*, **38**, 205 (1901); **61**, 129 (1907); Lewis and Randall, *J. Am. Chem. Soc.*, **43**, 233 (1921); *op. cit.*, Chaps. XXII and XXIII.

CHAPTER X

PHASE EQUILIBRIA

In the previous chapter the colligative, or osmotic, properties of dilute solutions containing solutes of low volatility were considered from the standpoint of thermodynamics; solutions of volatile substances, and concentrated solutions in general, are, however, preferably studied with the aid of the phase rule (p. 476) and the distribution law (p. 735), although it is sometimes convenient to employ Raoult's law and direct thermodynamic treatment. The general outlook in the present chapter is, consequently, somewhat different from that adopted in the one dealing with dilute solutions.

SOLUTIONS OF GASES IN LIQUIDS

Introduction.—The solubility of a gas in a liquid has a limited and definite value dependent on the nature of the gas and the liquid, and also on the temperature and pressure. For water as solvent, the most soluble gas known is ammonia, 1 ml. of water dissolving about 1300 ml. of the gas at 0° c. and 1 atm. pressure, whereas helium is the least soluble, dissolving to the extent of only 0.01 ml. in 1 ml. of water at the same temperature and pressure. The solubilities of other gases range themselves between these two extreme values. Increase of pressure results in an increase in the mass of gas dissolved, but increase of temperature has the opposite effect; the quantitative influence of these factors will be considered shortly. Qualitatively, the phase rule shows that both pressure and temperature are important. For a system consisting of a gas and a liquid solvent there are obviously two components and two phases, viz., gas and liquid, so that applying the phase rule

$$F = C - P + 2,$$

it follows that

$$F = 2 - 2 + 2 = 2,$$

and the system has two degrees of freedom; that is to say, for the state of the equilibrium to be completely defined two variables, i.e., pressure and temperature, must be fixed.

Determination of Solubility.—Various forms of apparatus have been used to measure the solubilities of gases in liquids, but the underlying principle is the same in most cases. For a sparingly soluble gas, some of the latter is enclosed over mercury in a gas-burette, e.g., of the type used in gas analysis, and its volume measured at a definite pressure and known temperature. The gas is passed at the same pressure into a vessel containing the gas-free solvent, and

the whole agitated in a suitable manner, at constant temperature, until the liquid is saturated; the residual gas is then transferred to the gas-burette and its volume measured. The difference between the original volume and the final value, after making a correction for the vapor pressure of the solvent, gives the amount dissolved in the quantity of solvent employed. For very soluble gases, e.g., ammonia and hydrogen chloride, this method is not suitable and an alternative, analytical, procedure is used. A thin-walled glass bulb, having a capacity of about 30 ml., with a bent inlet tube at the bottom and an outlet at the top, is weighed and then half-filled with solvent. A current of the gas is then bubbled through until the liquid is saturated. The ends of the inlet and outlet tubes are now sealed off by a flame, and the bulb and its contents are weighed. The difference between this and the previous weighing gives the weight of the solution. The quantity of solute contained therein is determined by analysis; since most very soluble gases are either acidic or basic, this presents no difficulty. If it is required to determine the solubility in unit volume of saturated solution, the density of the latter must be measured. The analytical procedure has been used to study the solubility of oxygen in water, for in spite of its relatively small concentration it can be readily estimated by Winkler's method.

Absorption and Solubility Coefficients.—Two coefficients have been employed to express the results of solubility measurements with gases: the first, proposed by R. Bunsen (1857), is known as the **absorption coefficient**. It may be defined as the volume of gas, reduced to 0° c. and 1 atm., dissolved by unit volume of solvent at the temperature of the experiment under a partial pressure of the gas of 1 atm. If v_0 is the volume of gas dissolved, reduced to s.t.p., V is the volume of solvent, p is the partial pressure of the gas in atm., then the absorption coefficient α is given by

$$\alpha = v_0/Vp. \quad (1)$$

The second coefficient, suggested by W. Ostwald (1888), is more useful in some ways: it is often called the **coefficient of solubility**, and is defined as the volume of gas, measured under the temperature and pressure at which the gas dissolves, taken up by unit volume of the liquid. If v is the actual volume of dissolved gas, measured at the experimental temperature T and at the partial pressure of the gas p atm., then assuming the gas laws to be obeyed,

$$v = Tv_0/T_0p, \quad (2)$$

where T_0 is 0° c., i.e., 273° K. According to definition the coefficient of solubility β is v/V , and combination of this with (1) and (2) gives

$$\beta = \alpha T/273. \quad (3)$$

It will be seen later that from certain points of view it is desirable to express solubilities in terms of mole fractions; in fact it would probably be better, from the theoretical standpoint, if this method were adopted universally. The coefficients defined above are nevertheless often used.

The absorption coefficients of a few of the common gases at 20° c. in water and in some organic solvents are quoted in Table 117.

TABLE 117. ABSORPTION COEFFICIENTS OF GASES AT 20° C.

	H ₂	He	N ₂	O ₂	CO	CO ₂	H ₂ S	NH ₃	HCl
Water	0.017	0.009	0.015	0.028	0.025	0.88	2.68	710	442
Carbon disulfide	0.031	—	0.049	—	0.076	0.83	—	—	—
Chloroform	—	—	0.120	0.205	0.177	3.45	—	—	—
Ethyl alcohol	0.080	0.028	0.130	0.143	0.177	3.0	—	—	—
Acetone	0.065	0.030	0.129	0.207	0.198	6.5	—	—	—
Ethyl ether	0.12	—	0.24	0.415	0.38	5.0	—	—	—
Benzene	0.066	0.018	0.104	0.163	0.153	—	—	—	—

Influence of Temperature.—When gases dissolve in water there is generally a liberation of heat; it follows, therefore, from the Le Chatelier principle (p. 831) that increase of temperature will result in a decrease of solubility. It is for this reason that gases may be readily expelled from solution by boiling. By thermodynamic methods it is possible to deduce the equation, for constant pressure, assuming the gas to be ideal,

$$\frac{d \ln c}{dT} = \frac{\Delta H}{RT^2}, \quad (4)$$

where c is the concentration in moles per liter of the gas in the liquid phase, and ΔH is the differential heat of solution of 1 mole of the gas in a saturated solution at the temperature $T^\circ \text{K}$. If ΔH may be assumed to be independent of temperature, then by integration of (4) it follows that

$$\ln \frac{c_2}{c_1} = -\frac{\Delta H}{R} \left(\frac{1}{T_2} - \frac{1}{T_1} \right), \quad (5)$$

where c_1 and c_2 are the concentrations of the solutions at T_1 and T_2 , respectively. Since according to Avogadro's law the number of moles of gas is proportional to its volume measured at s.t.p., it is evident that the concentration in moles per liter will be proportional to the Bunsen absorption coefficient α ; hence (5) may be written

$$\ln \frac{\alpha_2}{\alpha_1} = -\frac{\Delta H}{R} \left(\frac{1}{T_2} - \frac{1}{T_1} \right). \quad (6)$$

These equations are, of course, exactly analogous to the van't Hoff equation and its integrated form (p. 829); the concentration of the solution and the absorption coefficient are equivalent to the equilibrium constant of the system consisting of gas and saturated solution. It is also not difficult to see that these equations are related to the simplified form of the Clausius-Clapeyron equation (p. 452), the gas pressure being the equivalent of the vapor pressure as shown below.

Some values for the absorption coefficient of helium, nitrogen, oxygen and carbon dioxide at 0° and 30° c., in water as solvent, are quoted in Table 118. The relative influence of temperature is much greater for

TABLE 118. INFLUENCE OF TEMPERATURE ON SOLUBILITIES OF GASES IN WATER

	Helium	Absorption Coefficients		Carbon dioxide
		Nitrogen	Oxygen	
0° c.	0.0094	0.0235	0.0489	1.713
30°	0.0081	0.0134	0.0261	0.665

carbon dioxide than for the other gases; this means that the heat of solution is largest in this instance, which is no doubt related to the fact that chemical combination occurs between carbon dioxide and water.

Although it is generally true that increase of temperature is accompanied by a decrease of solubility of a gas, some sparingly soluble gases, such as hydrogen and the inert gases of the atmosphere, behave in an exceptional manner, especially in nonaqueous solvents. In water the solubility at first decreases with increasing temperature and then, after passing through a minimum, tends to increase; none of the effects is, however, very large. In organic solvents, e.g., hydrocarbons, alcohols and acetone, the solubilities of the inert gases show a slight increase with temperature, at least between 15° and 30° c. (A. Lannung, 1930).

Influence of Pressure: Henry's Law.—The most important factor influencing the solubility of a gas is pressure; as is to be expected from kinetic considerations, compression of the gas will tend to increase its solubility. The quantitative connection between solubility and pressure was first stated by W. Henry (1803); his results are included in the law bearing his name, that is **Henry's law**, which states that *the mass of gas dissolved by a given volume of solvent, at constant temperature, is proportional to the pressure of the gas with which it is in equilibrium*. If m is the mass of gas dissolved by unit volume of solvent at the equilibrium pressure p , then

$$m = kp, \quad (7)$$

where k is a constant. A corollary of Henry's law is that, provided the gas obeys Boyle's law, the volume of gas dissolved, *measured at the pressure of the experiment*, is independent of that pressure; this may be shown in the following manner. If the pressure of the gas is increased the mass dissolved will increase in the same proportion [equation (7)]. According to Boyle's law, however, the volume occupied by a given mass of gas is inversely proportional to the pressure, so that if the mass of gas is increased in proportion to the pressure, the volume must remain constant. It follows immediately from this form of Henry's law that the Ostwald solubility coefficient at a definite temperature should be independent of pressure; that this is very nearly true in many cases is to be regarded as confirmation of the law.

Henry's law can be expressed in still another way which provides a ready means whereby it may be verified. The mass m of gas dissolved per unit volume of solvent is really the concentration in g. per ml., and this is proportional to the concentration expressed in moles per liter of solvent,

or per liter of solution since there is no considerable volume change when the gas dissolves. Further, the pressure p of the gas is proportional to the number of molecules in the gas space; hence it is also proportional to the concentration in moles per liter of gas phase. According to Henry's law, equation (7), m/p is constant and therefore

$$\frac{\text{Concentration of gas in liquid phase}}{\text{Concentration of gas in gaseous phase}} = \text{constant} \quad (8)$$

at definite temperature, independent of pressure. This equation is a form of an important generalization known as the **distribution law**, which will be considered in more detail on p. 735. The results in Table 119, obtained by O. Sackur and O. Stern (1912), for the solubility of

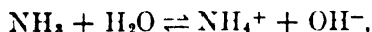
TABLE 119. DISTRIBUTION RATIO OF CARBON DIOXIDE IN VARIOUS SOLVENTS AT -59°C .

Equilibrium Pressure	Methyl Alcohol	Acetone	Methyl Acetate
100 mm.	42.5	67.2	75.8
200	42.7	68.0	77.1
400	43.1	69.2	77.6
700	43.3	72.8	79.0

carbon dioxide in a number of nonaqueous solvents at -59°C ., show that the **distribution ratio**, as given by (8), remains almost constant in spite of a seven-fold increase of pressure; it will be noted, however, that the pressures are all relatively low.

The result of many studies of the solubility of gases in liquids is to show that Henry's law is obeyed in a satisfactory manner for gases of low solubility, provided the pressures are not too high or the temperatures too low. Henry's law is, as shown above, an aspect of the general distribution law and also, as will be seen shortly, of Raoult's law; these generalizations are strictly applicable only to ideal systems, and any departure from ideal behavior of the gas will result in deviations from Henry's law. Such deviations are, therefore, to be expected at low temperatures and high pressures, especially with easily liquefiable gases.

Deviations from Henry's Law.—Other causes of departure from ideal behavior of the gas-liquid system are difference of molecular complexity in the two phases, compound formation between solute and solvent, and ionization. Ammonia, for example, shows considerable deviations from the requirements of Henry's law; this is especially true at low temperatures, and even at 100°C . the behavior is not ideal. It has been shown, however, that if allowance is made for the reaction



then Henry's law is obeyed at 100°C . (G. C'alingaert and F. E. Huggins, 1923). If c_l and c_g represent the total concentrations of the ammonia in the liquid and gaseous phases, respectively, then c_l/c_g is not constant; if α , however, is the fraction of the dissolved ammonia which has undergone

reaction with the water, then the ratio $c_1(1 - \alpha)/c_0$ is said to be constant, for the numerator now represents the concentration of free NH_3 molecules in the liquid phase. There are a number of complications in this problem, and it is possible that the constancy observed may be partly fortuitous.

The solubility of carbon dioxide in water at low temperatures and pressures of 10 to 100 atm. departs from Henry's law, but at 100°C . the results are in close agreement with it (W. Sander, 1912). The deviations at low temperatures are partly due to compound formation between carbon dioxide and water, but this cannot be the only factor responsible, since some nonaqueous solutions also show deviations; these may be attributed to the departure from ideal behavior due to other causes which would be appreciable at the high pressures employed.

It was discovered by W. D. Bancroft and B. C. Belden (1930) that the weight of hydrogen sulfide dissolved by unit volume of aniline is proportional to the pressure; in other words, the system obeys Henry's law. In view of the acidic nature of the solute and the basic properties of the solvent, compound formation would be anticipated with consequent departure from ideal behavior. The result observed may be due to the cancellation of disturbing factors operating in opposite directions.¹

Aqueous solutions of hydrogen chloride were found by H. E. Roscoe and W. D. Dittmar (1859) to depart very considerably from Henry's law; this is not surprising in view of the fact that the solution is a strong electrolyte which is almost completely dissociated into ions provided the concentration is not too high. It is probable, however, that the law does apply to the undissociated molecules of hydrogen chloride, although these constitute only a very minute proportion of the total amount dissolved. At high concentrations aqueous solutions of hydrogen chloride behave virtually as liquid-liquid systems, such as are discussed below, and Raoult's law appears to be applicable after making allowance for the removal of 0.25 mole fraction of the acid in the form of a nonvolatile compound with water (W. F. K. Wynne-Jones, 1930).

Solubility of Gas Mixtures.—If the gas being dissolved does not consist of a single chemical substance but is a mixture, the number of components of the system is increased and so also is the number of degrees of freedom. For a mixture of two gases, for example, C is 3, and from the phase rule it may be deduced that the system is trivariant; hence, for the composition of the liquid phase to be defined, it is necessary to fix the composition of the gas in addition to the temperature and pressure. The amount of a particular gas dissolved will thus depend on its proportion in the gaseous mixture. It was discovered by J. Dalton (1805) that in a mixture of gases each constituent dissolves according to its own partial pressure (cf. p. 302); in other words, Henry's law applies to each gas independently of the pressure of the others. For example, in dry air free from carbon dioxide at a total pressure of 1 atm., the partial pressures of the chief gases, nitrogen, oxygen and argon, are 0.7806, 0.2100 and 0.0094 respectively. If m_1 , m_2 and m_3 are the masses of these elements

which dissolve in unit volume of water at 1 atm. pressure and at a definite temperature, e.g., 0° c., the amounts dissolved from air will be in the following ratio:

$$\begin{array}{ccc} \text{N}_2 & \text{O}_2 & \text{A} \\ 0.7806 m_1 & 0.2100 m_2 & 0.0094 m_3. \end{array}$$

It will be evident, from the definition on p. 694, that the Bunsen absorption coefficient is proportional to the quantity m ; the coefficients for the three gases are 0.0235, 0.0489 and 0.0578, respectively, and the volumes of the gases dissolved from a large quantity of air, so that the partial pressures remain almost unchanged, will be in the following proportions:

$$\begin{array}{ccc} \text{N}_2 & \text{O}_2 & \text{A} \\ 0.7806 \times 0.0235 & 0.2100 \times 0.0489 & 0.0094 \times 0.0578 \\ \text{i.e.} & 0.01834 & : \quad 0.01027 \quad : \quad 0.000543. \end{array}$$

From these results it appears that the dissolved gas should contain 35.2 per cent by volume of oxygen; the actual value, with air at 0° c. and 1 atm. pressure, obtained by L. W. Winkler (1891) was 34.9 per cent. As the solubility coefficient of oxygen is higher than that of nitrogen, the gas expelled when water saturated with air is heated contains more of the former constituent than does the original air. This procedure of the solution of air followed by expulsion, the process being repeated several times, was at one time proposed as a method for the extraction of oxygen from the atmosphere. It can be easily calculated that after six successive absorptions and expulsions of the air from water, the final gas expelled should contain over 90 per cent of oxygen by volume. The foregoing treatment is based on the use of Dalton's law and Henry's law, and hence is strictly applicable to an ideal gas mixture only. For real gases deviations would be observed, especially at appreciable pressures (cf. p. 305).

Solubility of Gases in Aqueous Solutions.—As a general rule gases are less soluble in aqueous solutions of electrolytes than in pure water; this is an example of the phenomenon known as the **salting-out effect**. The relative influence of a given salt is almost independent of the nature of the gas; the ratios of the solubilities (s) of some gases in 1.0 N solutions of lithium chloride and of sodium sulfate to the values (s_0) in pure water are given in Table 120. It appears that sodium sulfate reduces the

TABLE 120. THE SALTING-OUT OF GASES FROM AQUEOUS SOLUTION

Solution	Hydrogen	Nitrogen	Ratio s/s_0 Carbon Dioxide	Hydrogen Sulfide
1.0 N Lithium chloride	0.84	0.79	0.83	-
1.0 N Sodium sulfate	0.73	0.70	0.68	0.73

solubility of a gas to a greater extent than does lithium chloride. From a consideration of the influence of a number of salts it may be concluded that the salting-out effect of an ion is larger the smaller its size and the greater the charge it carries. The variation of gas solubility with the concentra-

tion of the salt solution is represented, at least approximately, by an equation of the form

$$\log s/s_0 = kc, \quad (9)$$

where c is the concentration of the salt solution and k is a constant for the given salt.

The salting-out effect of an electrolyte has been ascribed to the congregation of dipolar molecules of the polarizable solvent, e.g., water, around the ions of the added electrolyte. If the solute is only slightly polarizable and either nonpolar or only slightly polar, as is the case with the less soluble gases, the molecules will not be attracted by the ions. The effect will then be similar to that produced by molecular forces which result in positive deviations from Raoult's law (p. 677); the escaping tendency of the solute is thus increased and hence its solubility is diminished. Nonelectrolytes, such as sugar, also have the effect of reducing the solubility of gases in aqueous solution; the diminution is quite considerable in spite of the absence of ionic forces. The decrease in solubility may nevertheless be due to a similar cause, namely, the preferential orientation of the solvent molecules around those of the added substance, for the latter is usually polar even if it is not ionized.

Attempts have been made to calculate the "hydration numbers" i.e., the extent of hydration, of various ions and nonelectrolytes from gas solubility measurements. Assuming the fundamental solubility of the gas to remain unaffected by the added substance, it is possible to calculate the amount of water in solution that is "free"; the remainder is supposed to be bound by the added ions or molecules, and so the apparent extent of hydration can be evaluated. The hydration numbers obtained in this manner are in satisfactory agreement with those estimated in other ways (cf. p. 920); too much emphasis, however, should not be laid on this agreement as it may be fortuitous.²

Henry's Law and Raoult's Law.—Instead of representing Henry's law by (7) as $m = kp$, m being the mass of gas dissolved in unit volume of solvent, it may be transformed so as to give the mass of gas dissolved by unit weight of solvent; thus,

$$\frac{w_2}{w_1} = k'p_2, \quad (10)$$

where w_1 is the weight of solvent, w_2 is the weight of gas dissolved in it, and p_2 is the pressure of the gas. Dividing the two masses by the respective molecular weights, it follows that

$$\frac{n_2}{n_1} = k''p_2, \quad (11)$$

where n_1 and n_2 are now the numbers of moles of solvent and solute, respectively. For a dilute solution, that is, for a gas that is not too soluble, n_2 is small in comparison with n_1 , so that $n_1 + n_2$ may replace n_1 in the denominator; hence,

$$\frac{n_2}{n_1 + n_2} = x_2 = k''p_2, \quad (12)$$

where x_2 is the mole fraction of gaseous solute in the saturated solution. The solubility of a gas, expressed in mole fractions, is thus proportional to the pressure of the gas; this may be regarded as the ideal form of Henry's law.

The solubility of a gas can be considered from two points of view: x_2 may be regarded as the solubility of the gas, in mole fractions, under the pressure p_2 , or alternatively p_2 may be taken as the vapor pressure of a volatile solute when it is present to the extent of mole fraction x_2 in the solution. Hence Henry's law may be stated in the following form: *in a dilute solution the vapor pressure of a solute is proportional to its mole fraction*. In the special case in which the law is applicable over the whole range of concentrations, from an infinitely dilute solution, i.e., pure solvent, to pure (liquid) solute, then p_2 in (12) becomes p_2^0 , the vapor pressure of pure solute, when x_2 is unity; hence,

$$1 = k''p_2^0 \quad \text{or} \quad k'' = 1/p_2^0 \quad (13)$$

and so (12) may be written

$$x_2 = p_2/p_2^0 \quad \text{or} \quad p_2 = x_2p_2^0, \quad (14)$$

identical with the form Raoult's law takes for a volatile solute (p. 710). It is seen, therefore, that Raoult's law may be regarded as a special case of Henry's law; all systems which obey the former must satisfy the latter, but the reverse will only be true if Henry's law applies over the whole range of concentrations.

It has been tacitly assumed that the solvent is nonvolatile, but this is not necessarily the case. The volatility of the solvent is quite independent of that of the solute; in fact if the ideal form of Henry's law applies to the latter, then it can be shown that Raoult's law must apply to the vapor pressure of the solvent. If the vapors behave as ideal gases, and this is approximately true because of the low pressures, then the Duhem-Margules equation, deduced on p. 708, may be applied in the form

$$x_1 \frac{d \log p_1}{dx_1} - x_2 \frac{d \log p_2}{dx_2} = 0, \quad (15)$$

$$\therefore \frac{d \log p_1}{d \log x_1} - \frac{d \log p_2}{d \log x_2} = 0, \quad (16)$$

where x_1 and x_2 refer to the mole fractions, and p_1 and p_2 to the vapor pressures of the solvent and solute, respectively. Taking logarithms of (12) there results

$$\log x_2 = \log k'' + \log p_2,$$

and differentiation gives

$$\frac{d \log p_2}{d \log x_2} = 1, \quad (17)$$

and hence, from (16),

$$\frac{d \log p_1}{d \log x_1} = 1. \quad (18)$$

Upon integrating and extracting the logarithms, it follows at once that

$$p_1 = kx_1. \quad (19)$$

In this equation the constant k is equal to p_1^0 , the vapor pressure of the pure solvent, as may be seen by putting x_1 equal to unity when p_1 becomes p_1^0 ; hence (19) is equivalent to

$$p_1 = x_1 p_1^0, \quad (20)$$

and this is Raoult's law for the solvent (p. 627). From this argument it is evident that in any solution Raoult's law is applicable to the solvent over the same range that Henry's law, in the form of (12), applies to the solute.

Ideal Solubility and Nature of the Gas.—It is possible by means of (14) to calculate the solubility of a gas below its critical temperature, on the assumption that the solution behaves in an ideal manner. The critical temperature of ethane is 34°C ., and at 25°C . the pure liquid has a vapor pressure (p_2^0) of 42 atm.; according to (14), therefore, the solubility of ethane at 25°C . and a pressure of 1 atm. in any solvent is given, in mole fractions, by the relationship

$$x_2 = \frac{1}{42} = 0.024 \text{ mole fraction,}$$

since p_2 is 1, and p_2^0 is 42 atm. The actual solubility in *n*-hexane at 25°C . and 1 atm. pressure is 0.017 mole fraction; the difference between this and the calculated value shows that the system is not ideal.

In order to extend the method for calculating gaseous solubilities to temperatures above the critical point, it is necessary to estimate the hypothetical vapor pressure of the liquid by a suitable extrapolation; this is best done with the aid of the integrated Clapeyron-Clausius equation (p. 453). If the vapor pressure at any two temperatures is known, the value at any other temperature may be evaluated on the assumption that the molar heat of vaporization remains constant. The critical temperature of methane is -95.5°C ., and the hypothetical vapor pressure of the liquid at 20°C . may be extrapolated as approximately 310 atm., giving an ideal solubility at this temperature and a pressure of 1 atm. of $1/310 = 0.0032$ mole fraction; this is very close to the solubilities actually found in *n*-hexane and in *m*-xylene. Since the solubility in mole fractions of a gas at 1 atm. pressure is equal to $1/p_2^0$, where p_2^0 is the vapor pressure of the liquefied gas, it is evident that, for ideal solutions, the lower the vapor pressure at the given temperature the greater will be the solubility of the gas. Gases which are liquefied only with difficulty, that is to say, those having very low boiling points, may be regarded as having high vapor pressures; such gases will, therefore, have low solubilities. It follows, therefore, that in general, easily liquefiable gases will be the most soluble; this is in agreement with observation in most cases.

Although the solubility of a gas, in mole fractions, should theoretically be independent of the nature of the solvent, this is not true in practice because of departure from ideal behavior. The solubilities of the non-polar gases nitrogen, oxygen, carbon monoxide and argon, and probably

also of hydrogen and helium, are of the right order, viz., from 50 to 100 per cent, of the ideal value in various solvents, such as hydrocarbons, carbon tetrachloride, and some simple esters. These substances are not only nonpolar and nonassociated, but they also have low internal pressures of the same order as those of the liquefied gases. In polar solvents of relatively high internal pressure, e.g., nitrobenzene, alcohols and water, the solubilities are much lower than the ideal values, for these systems exhibit deviations from ideal behavior (p. 677). Some data for solutions of gases, showing the extent of the deviations, are recorded in Table 121. The solubilities in water are exceptionally low, but this is not surprising

TABLE 121. IDEAL AND OBSERVED SOLUBILITIES AT 20° C.

Gas	Mole Fractions $\times 10^4$				Water
	Ideal	Nitrobenzene	Ethyl Alcohol	Aniline	
Nitrogen	10	2.6	3.3	1.1	0.13
Carbon dioxide	11	3.9	4.5	1.9	0.19
Oxygen	16	—	—	—	0.17
Argon	21	—	6.5	—	0.41

since water is both polar and associated, and also has a very high internal pressure; solutions of gases of the type mentioned in the table would hardly be expected to behave ideally. Even chlorine and carbon dioxide, which interact with water and are generally regarded as relatively soluble gases, have solubilities considerably below the calculated values, because of their low polarity and internal pressure. A quite different type of behavior is shown by ammonia, which is a highly polar substance with a high internal pressure. In hydrocarbon solvents, therefore, its solubility is considerably below the ideal value, whereas in alcohol and in water the observed solubility is somewhat greater than that calculated. If allowance could be made for interaction between ammonia and the solvent, as mentioned above (p. 697), good agreement would probably be found. A corollary to the foregoing conclusions is that for a number of gases of similar polarity and internal pressure, e.g., hydrogen, nitrogen, carbon monoxide, oxygen and the inert gases, which do not react with the solvent, the ratio of the solubilities in various solvents should be approximately independent of the nature of the gas; this generalization is roughly true in practice, and only gases such as carbon dioxide and ammonia, which are not in the same category as those mentioned above, are exceptional.³

BINARY LIQUID SYSTEMS

Completely Miscible Liquids.—Mixtures consisting of two volatile liquids, which are completely miscible with one another in all proportions, have a great deal in common with the gas-liquid systems already discussed. It can be seen that there is a fundamental similarity between a solution of a gas in a volatile solvent, e.g., oxygen in alcohol, and a mixture of two volatile liquids, e.g., ether and alcohol; in view of the low

solubility of the gas, however, its vapor pressure will be much higher than that of ether in a solution of equal mole fraction. The phase rule applied to a binary system of completely miscible liquids in equilibrium with vapor, i.e., two components and two phases, again shows that there are two degrees of freedom. The system will, therefore, be defined completely if two variables are fixed. In the study of liquid-vapor equilibria the emphasis is laid on the measurement of vapor pressure for liquid mixtures of definite composition, instead of fixing the pressure and determining the composition of the mixture, as is done in the case of gas-liquid equilibria. The two variables chosen are consequently temperature and the composition of the liquid, and if these are fixed the system must have a definite vapor pressure. Since the vapor contains both constituents of the mixture, the phase rule indicates that the composition of the vapor, as well as its pressure, must be definite; that is to say, the partial pressures of both constituents, as well as the total vapor pressure, are determined by the composition of the liquid and its temperature. It is to be expected, therefore, that there will be some relationship between the partial pressures of the two components in the vapor phase.

Measurement of Vapor Pressures.—In order to determine the partial vapor pressures for a mixture of two liquids, the method generally employed is to measure the total vapor pressure and the composition of the vapor in equilibrium with a given liquid mixture at a definite temperature; by assuming Dalton's law of partial pressures the individual pressures of the two constituents can then be calculated. For example, if the total pressure is found to be P , and analysis of the vapor shows it to contain n_A moles of component A to n_B moles of B,* then the respective partial pressures p_A and p_B are given by

$$p_A + p_B = P, \quad (21)$$

$$p_A = \frac{n_A}{n_A + n_B} P \quad \text{and} \quad p_B = \frac{n_B}{n_A + n_B} P, \quad (22)$$

according to Dalton's law. The procedures adopted for determining the various quantities required fall into three categories similar to those considered in connection with the vapor pressures of pure liquids and of solutions of nonvolatile solutes.

(i) **Static Methods.**—The simplest method is to place a mixture of known composition in an evacuated vessel and to measure the total vapor pressure on a suitable manometer when equilibrium is attained at a definite temperature (cf. p. 446); a quantity of the vapor is then removed and analyzed. In order not to change appreciably the composition of the liquid and to insure that the vapor removed should be in equilibrium with it, attempts have been made to carry out the analyses on very small amounts of vapor; this has been achieved

* In the study of miscible liquids it is not possible, nor is it desirable, to distinguish between solvent and solute; the designations A and B are therefore employed instead of 1 and 2, as in Chapter IX.

by measuring the refractive index of the vapor by an interferometer. The principle has been used by a number of workers, but the results may not be very accurate because of the uncertainty that liquid-vapor equilibrium has been attained. This difficulty has been overcome in the apparatus employed by J. B. Ferguson and his collaborators (1929 *et seq.*). A known weight of liquid of definite composition is distilled completely into the evacuated apparatus, using liquid air as the refrigerant; the liquid under examination is then transferred to the "reaction chamber," and placed in a thermostat, where vaporization occurs. By means of a reciprocating-pump device the considerable volume of vapor produced is made to pass back and forth over the liquid for some hours, thus insuring complete attainment of equilibrium. When this condition is reached, as indicated by the constancy of pressure, the total vapor pressure is read on a manometer. A definite volume of the vapor is now trapped by means of mercury, and completely condensed into a bulb, which can be sealed off, removed and the weight of its contents determined. From the data obtained it is possible to calculate the compositions of liquid and vapor in equilibrium in various ways; the description of one method will, however, be sufficient here. From the volume, weight, temperature and pressure of the vapor it is possible to calculate its composition on the assumption that it behaves as an ideal gas, since the molecular weights of the two constituents are known; the actual densities of the separate vapors can, if required, be determined in the same apparatus and the assumption made that the simple mixture law applies. From the composition of the vapor and its total pressure the partial pressures of the two components can be calculated. Knowing the total weights of the substances introduced into the apparatus, and the amounts in the vapor having been determined, the composition of the residual liquid in equilibrium with the vapor is readily evaluated.

An interesting modification of the static method has been employed by G. Calingaert and L. B. Hitchcock (1927). A mixture of known amounts, e.g., n_A and n_B moles, of the two components A and B is introduced over mercury into a graduated tube similar to a nitrometer in construction; by lowering the level of the mercury the volume is increased, so that vapor is produced and equilibrium is attained between vapor and liquid. The volume of vapor is measured and so also is its pressure, by means of a manometer attached to the nitrometer tube. Since the temperature is known the total number of moles (N_v) in the vapor can be calculated, on the assumption that the gas laws are obeyed; this assumption is shown to result in only small errors. Let x_l represent the mole fraction of A in the liquid phase, and x_v the value in the vapor phase; then the two equations readily follow:

$$(i) \quad N_l + N_v = n_A + n_B \quad \text{and} \quad (ii) \quad N_l x_l + N_v x_v = n_A, \quad (23)$$

where N_l , the total number of moles in the liquid phase, can be obtained directly from the first of these equations, for the other three quantities are known. It is, however, still necessary to determine x_l and x_v , and in order to do so application is made of the fact, based on the phase rule, that provided the temperature and pressure are fixed, the composition of liquid and vapor in equilibrium will be definite and independent of the *amounts* of the two phases. Starting with the two components in different proportions, represented by n'_A and n'_B , so that the total number of moles in liquid and vapor phases are now N'_l and N'_v , then provided the temperature and vapor pressure are the same as in the

previous case, the composition of liquid and vapor, and hence x_l and x_v , are unchanged; therefore,

$$N'_l x_l + N'_v x_v = n'_A. \quad (24)$$

There are then two simultaneous equations, (23, ii) and (24), available from which x_l and x_v can be deduced, thus giving the compositions of the liquid and vapor in equilibrium. Since the total vapor pressure is known the partial pressures can be calculated as described above.⁴

(ii) **Dynamic Methods.**—In the dynamic method the liquid is made to boil by adjusting the external pressure at a definite temperature; the vapor is then condensed and analyzed. The first reliable determinations of partial pressures of binary liquid systems were made by J. von Zawidski (1900), who used this method; it has been subsequently improved in a number of ways to insure equilibrium between liquid and vapor. The modern form of the apparatus is based on that described by J. Sameshima (1918); its principle can be explained by means of the diagram in Fig. 158. The vessel *A*, about 200 cc. capacity, contains the experimental liquid; into the liquid dips a thermometer and an

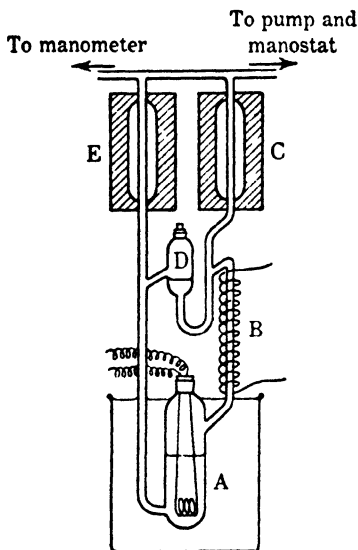


FIG. 158. Partial vapor pressures by the dynamic method

electrical heating-coil of fine wire, the object of which is to aid the evolution of bubbles and to prevent superheating (cf. p. 638). The apparatus is immersed in a thermostat as shown. The pressure is adjusted by a pump and a manostat so that the liquid in *A* boils gently at the temperature at which it is desired to measure the vapor pressure. The vapor ascends tube *B*, surrounded by a heating-coil to prevent condensation, is completely liquefied in *C*, which should be a very efficient condenser, and the resulting liquid runs down into the vessel *D*, the overflow returning to *A*. Any vapor coming from the liquid in *D* is condensed in *E*, and also returns to the boiling tube. The process of gentle ebullition, with distillation and overflow, is allowed to continue for some time, e.g., 30 to 40 minutes, until a steady state is attained, when the pressure on the manometer is read. The liquid collected in *D* has the same composition as the vapor from which it has condensed, and this is in equilibrium with the boiling liquid. Samples of liquid in *A* and *D* are rapidly with-

drawn, the former is cooled to prevent evaporation, and then analyzed; physical methods, involving measurement of density or refractive index, are generally employed. All the data are now available for calculating the partial pressures of the two constituents for one composition of the liquid mixture. The process is repeated with a number of other liquid mixtures.⁵

(iii) **Transpiration Methods.**—Some of the earliest applications of the transpiration or gas-saturation method to the study of liquid mixtures were made by C. E. Linebarger (1895) and R. A. R. Gahl (1900). The principle is identical with that already described for pure liquids (p. 448) and for solutions

(p. 630); it is applied in two ways which differ slightly. In the first method, dry air, or other gas, e.g., nitrogen or hydrogen, is drawn first through a saturator containing one of the constituents A, of the mixture A and B, in the pure state, then through a suitable absorber; if a chemical absorber is not available, then physical methods, e.g., condensation by liquid air, or adsorption by charcoal or silica gel, may have to be employed. The gas next passes into a pre-saturator, then on through the liquid mixture under examination, and finally all the vapor is removed in another absorber. It is essential that the amounts of the constituents taken up by the latter should be known separately; if the vapor is merely condensed then its composition could be determined by density or refractive index. Alternatively, two separate chemical absorbers, each taking up one constituent only, could be employed. If the two saturators are referred to as I (pure A) and II (mixture of A and B) it can be readily deduced, on the assumption that the ideal gas laws and Dalton's law are applicable (cf. p. 302), that

$$\frac{v_I}{v_{II}} = \frac{B_{II} - p_A - p_B}{B_I - p_A^0}, \quad (25)$$

where v_I and v_{II} are the volumes of dry air, or other gas, passing through the two saturators; B_I and B_{II} are the total pressures in them, measured by manometers, p_A and p_B are the partial pressures of the two constituents in II, and p_A^0 is the vapor pressure of the pure A in I. If n_A^0 moles of A are taken up in the first absorber, and n_A and n_B moles of A and B, respectively, are absorbed at the end, then

$$\frac{n_A^0}{n_A} = \frac{p_A^0 v_I}{p_A v_{II}} \quad \text{and} \quad \frac{n_A^0}{n_B} = \frac{p_A^0 v_I}{p_B v_{II}}, \quad (26)$$

$$\therefore p_A = p_A^0 \frac{n_A^0 (B_{II} - p_A - p_B)}{n_A (B_I - p_A^0)} \quad (27)$$

and

$$p_B = \frac{n_B}{n_A} p_A. \quad (28)$$

To evaluate p_A it is necessary that p_A^0 , the vapor pressure of the pure liquid A, should be known at the temperature of the saturator. If it is not available then a separate experiment must be carried out with pure A and its vapor pressure determined by the method described on p. 448, for which the volume of air or other gas passing through the saturators must be known.

The alternative method involves the use of only one saturator, preceded by a pre-saturator, which contains the liquid mixture being studied, and followed by an absorber permitting the amount of each constituent to be determined, as mentioned above. The volume of gas passing through the apparatus is measured, and the calculation of the total pressure is made in the ordinary manner as for a pure liquid (p. 448). As the composition of the vapor is known from the analysis of the material absorbed at the end of the chain, the partial pressure of each constituent can be calculated. The air-saturation methods are probably most useful for measurements at room temperature, or thereabouts, whereas for higher temperatures the dynamic method may be preferred.⁶

The Duhem-Margules Equation.—Before discussing the results of experimental determinations of vapor pressures of mixtures, certain

theoretical aspects must be considered. It was seen above that application of the phase rule leads to the conclusion that the partial vapor pressures of the two components will be related in some manner; this relationship can be deduced by simple thermodynamic methods. Since the chemical potential μ is the partial molar free energy, the Gibbs-Duhem equation for a system of two components A and B can be written in the form (cf. p. 238)

$$n_A d\mu_A = -n_B d\mu_B, \quad (29)$$

where n_A and n_B are the numbers of moles of the two constituents in a given mixture. Dividing both sides by $n_A + n_B$, there is obtained

$$x_A d\mu_A = -x_B d\mu_B, \quad (30)$$

where x_A and x_B are the respective mole fractions. Because of various mathematical considerations, since μ_A and μ_B are both functions of x_A , remembering that x_B is equivalent to $1 - x_A$, it is possible to convert (30) into

$$x_A \frac{d\mu_A}{dx_A} = -x_B \frac{d\mu_B}{dx_A} = x_B \frac{d\mu_B}{dx_B}, \quad (31)$$

since dx_A is equivalent to $-dx_B$.

The chemical potential of any constituent of a liquid mixture is given by (p. 683)

$$\mu = \mu^0 + RT \ln p, \quad (32)$$

where p is the partial pressure of that constituent, the vapor being assumed to behave as an ideal gas; more correctly fugacities should be used, but as the partial pressures are low, the difference can be neglected. Differentiating (32) with respect to x_A and x_B , respectively, at constant temperature and *constant total pressure*, the results are

$$\frac{d\mu_A}{dx_A} = RT \frac{d \ln p_A}{dx_A}, \quad (33)$$

$$\frac{d\mu_B}{dx_B} = RT \frac{d \ln p_B}{dx_B}. \quad (34)$$

Combination with (31) then gives

$$x_A \frac{d \ln p_A}{dx_A} = x_B \frac{d \ln p_B}{dx_B}, \quad (35)$$

$$\therefore \frac{d \ln p_A}{d \ln x_A} = \frac{d \ln p_B}{d \ln x_B}. \quad (36)$$

Equations (35) and (36) are alternative forms of what is generally called the **Duhem-Margules equation**; it was first derived by J. Willard Gibbs (1876), and later, independently, by P. Duhem (1886), M. Margules (1895) and R. A. Lehfeldt (1895). It shows that there must be some

connection between the relative amounts of the two constituents and p_A and p_B , their partial vapor pressures.*

The actual relationship between p_A and p_B cannot be derived from (35) or (36) without making an arbitrary assumption which permits the equation to be integrated; the integration assumed by M. Margules (1895) has been put into the simple form

$$(a) \quad p_A = p_A^0 x_A e^{\alpha x_B^2} \quad \text{and} \quad (b) \quad p_B = p_B^0 x_B e^{\alpha x_A^2}, \quad (37)$$

where α must be the same constant for both substances, and p_A^0 and p_B^0 are, as usual, the vapor pressures of the pure constituents. Equations of this type appear to describe in a satisfactory manner the variation of the partial pressure of each constituent of a liquid mixture with the composition of the latter, at constant temperature, except in such cases as when one or other of the components is highly associated.⁷

If the partial pressure p_A of one constituent of a mixture is measured, then since p_A^0 is known, α may be calculated from (37a), and hence p_B determined by (37b) if p_B^0 is also known; thus the two partial pressures are seen to be directly related. An alternative possibility is to calculate p_A and p_B from a knowledge of the total pressure P ; since the latter is equal to the sum of the two partial pressures,

$$P = p_A + p_B, \quad (38)$$

$$\therefore \frac{dP}{dx_A} = \frac{dp_A}{dx_A} + \frac{dp_B}{dx_A}. \quad (39)$$

Differentiating (37a) and (37b) with respect to x_A , remembering that $x_A + x_B = 1$, and adding, an expression for dP/dx_A is obtained; two special cases may be considered, viz.,

$$(a) \quad x_A = 0, \quad \left(\frac{dP}{dx_A} \right)_{x_A=0} = p_A^0 e^{\alpha} - p_B^0 \quad (40)$$

and

$$(b) \quad x_A = 1, \quad \left(\frac{dP}{dx_A} \right)_{x_A=1} = p_A^0 - p_B^0 e^{\alpha}. \quad (41)$$

If the value of dP/dx_A is obtained graphically, from the slope of the total pressure-composition curve at either $x_A = 0$ or $x_A = 1$, the constant α can be evaluated, since p_A^0 and p_B^0 are known for the pure substances. When the vapor pressure curves are of a more complicated type it may be necessary to introduce an additional term, of the type βx^3 , into the exponential parts of equations (37a) and (37b); the two constants α and β can then be determined by solving simultaneous equations for $x_A = 0$ and $x_A = 1$, both of which contain α and β .

Ideal Systems: Raoult's Law.—In the study of liquid mixtures it is convenient to define an ideal solution as one obeying Raoult's law over

* The only assumption involved in the deduction of the Duhem-Margules equation is that the vapors behave ideally; it should be noted that no assumption is involved as to the ideality, or otherwise, of the liquid mixture.

the whole range of concentration; as has been seen in Chapter IX there are satisfactory theoretical reasons for adopting this definition. It can be readily shown that if one constituent of a mixture obeys the ideal law, so also must the other; this is exactly analogous to the conclusion reached concerning gas-liquid systems, that if Raoult's law applies to the liquid then Henry's law must hold for the gas (p. 701). Suppose Raoult's law, that the partial vapor pressure of a constituent is proportional to its mole fraction in the liquid *at all compositions*, applies to A, then

$$p_A = p_A^0 x_A, \quad (42)$$

and on taking logarithms and differentiating with respect to x_A it follows that

$$x_A \frac{d \ln p_A}{dx_A} = 1. \quad (43)$$

According to equation (35), which applies whether the solution is ideal or not, it is seen that the analogous expression involving B must also equal unity; that is,

$$x_B \frac{d \ln p_B}{dx_B} = 1, \quad (44)$$

and integration, remembering that p_B becomes p_B^0 when x_B is 1, gives

$$p_B = p_B^0 x_B, \quad (45)$$

which is Raoult's law for the component B. The same conclusion may be reached by putting α equal to zero in equations (37a and b) which

simultaneously satisfy the Duhem-Margules rule; it is seen that (37a) reduces to (42), and (37b) to (45), so that Raoult's law is applicable to both constituents of the mixture. If α is not zero then the ideal law applies to neither.

If a mixture of two liquids behaves ideally, then according to the equations $p_A = p_A^0 x_A$ and $p_B = p_B^0 x_B$, the plot of the partial pressure of each constituent against its respective mole fraction in the liquid phase should be a straight line passing through the origin, as shown by the thin lines in Fig. 159. The total vapor pressure, i.e., the sum of the two

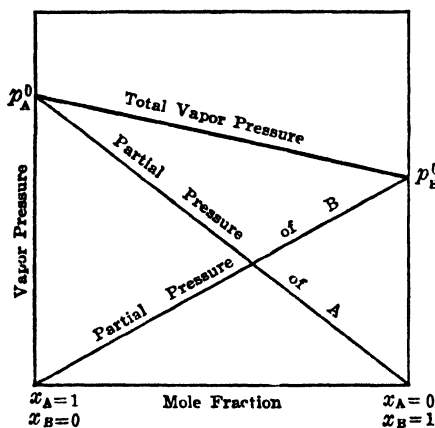


FIG. 159. System obeying Raoult's law

partial pressures, is then also a straight line, as indicated in the diagram. Only a limited number of cases are known of mixtures which obey Raoult's

law over the whole range of concentration and give straight-line plots of vapor pressure against the composition of the liquid expressed in mole fractions; some examples are the following: ethylene dibromide and propylene dibromide at 85°; benzene and ethylene dichloride at 50°; *n*-hexane and *n*-heptane at 30°; ethyl bromide and ethyl iodide at 30°; and *n*-butyl chloride and *n*-butyl bromide at 50° c. A number of other cases have been reported in the literature, but the experimental evidence is not always free from objection. The temperatures mentioned are those at which measurements have been made; it is probable that similar results would be obtained at other temperatures, but this has not yet been established. It will be noted that in all except one of the systems mentioned as obeying Raoult's law the two components are very closely similar, physically and chemically; it is just such mixtures that would be expected to behave ideally, for the intermolecular forces of the two components are almost identical and there is little difference in the environment in the pure liquid and the mixture (cf. p. 676). The ideal behavior of the benzene-ethylene dichloride system is perhaps surprising, and it may be due to cancellation of two effects acting in opposite directions; this could be verified by experiments over a range of temperature.⁸

Positive Deviations.—It was shown in Chapter IX that if the two constituents of a mixture differ in (a) internal pressure, (b) polarity, (c) length of hydrocarbon chain or analogous grouping, or if (d) one or other of the substances is associated in the liquid state, the system does not behave ideally; the four factors mentioned all result in positive deviations from Raoult's law. The measured vapor pressures are greater than those to be expected from equations (42) and (45), and are in harmony with equations (37a) and (37b) with α having a positive value. It is evident from these relationships, which satisfy the Duhem-Margules equation, that if one component of a mixture shows positive deviations from Raoult's law so also must the other. Systems with positive deviations are, as is to be anticipated, very common; in fact, most homogeneous liquid mixtures fall into this category. In some cases the deviations are very small, e.g., carbon tetrachloride and heptane at 50° c., both constituents being nonpolar and of low internal pressure; in others they are larger, e.g., ethyl ether and acetone, at 20° and 30° c., the components differing appreciably in internal pressure; whereas in other instances the deviations are so considerable, e.g., heptane and ethyl alcohol at 50° c., that the partial pressures seem to bear no relationship to those expected from Raoult's law. In the last mentioned system not only do the two components differ markedly in internal pressure, polarity and chain lengths, but the alcohol is highly associated in the liquid state; all four factors leading to positive deviations from ideal behavior are therefore present.

The type of vapor pressure curves obtained for mixtures of the kind under consideration is shown in Fig. 160, where the dotted lines indicate

ideal behavior; the experimental curves generally approach the ideal ones as the system approaches pure A or pure B. This means that the "solvent," e.g., A, tends to obey Raoult's law when the concentration of the "solute," i.e., B, becomes very small. Similarly, B satisfies Raoult's law when the proportion of A is small. At the same time, that is, at low concentrations, the "solute" must in each case obey Henry's law (cf. p. 701); the slope of the partial pressure curve in Fig. 160 must then be

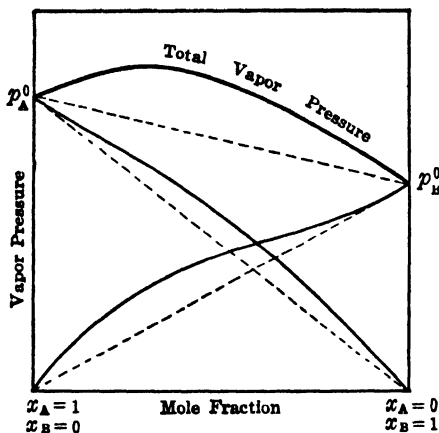


FIG. 160. Positive deviations from Raoult's law

linear, but not necessarily the same as for a completely ideal solution, i.e., the broken line. Solutions of this kind are the ideal dilute solutions referred to on pp. 675, 685.

If the positive deviations from ideality are large, the total vapor pressure curve may show a maximum; such curves are quite common, as will be seen shortly. It is obvious that if the two components of a binary system have vapor pressures that are almost the same, the theoretical total vapor pressure for an ideal system should be a horizontal line. Any positive deviations, even if quite small, must then result in a

curve having a maximum, indicating that there is a maximum vapor pressure for one particular composition of the liquid mixture.

When a system consists of two substances which differ considerably, i.e., all the properties favor positive deviations, and especially if one of them is associated, there is a possibility that the components will become incompletely miscible as the temperature is lowered. Where there is a tendency for this to occur the partial vapor pressure curves often have a relatively flat portion; the total pressure curve then shows a flat maximum extending over a considerable range of concentration. This type of behavior, e.g., with *n*-pentane and ethyl alcohol, generally indicates that the homogeneous liquid will separate into two layers as the temperature is diminished. It may be mentioned that ethyl alcohol and *n*-hexane, with a hydrocarbon chain slightly longer than that of pentane, are only partially miscible, and form two layers, at ordinary temperatures.

The effect of increasing hydrocarbon chain length in favoring positive deviations is well brought out by considering mixtures of water with a primary alcohol. When the latter is methyl alcohol the total pressure increases steadily from pure water to the pure alcohol, but with ethyl alcohol it rises rapidly at first and then more slowly with a very slight maxi-

mum. Mixtures of water and *n*-propyl and especially isobutyl alcohol have vapor pressure curves with broad and flat maxima.

Negative Deviations.—If the two constituents of a mixture are such that the molecules of A and B attract one another strongly, and particularly if there is partial compound formation between A and B in the liquid phase, the vapor pressure of each constituent may be less than that required by Raoult's law. Such a system can, therefore, exhibit negative deviations from ideal behavior. The influence of the attraction between A and B must, however, outweigh the effects of the differences mentioned above which are accompanied by positive deviations; it is not surprising, therefore, that cases of negative deviation from ideal behavior are uncommon. In such instances α in equations (37a and b) is negative, and both components have lower vapor pressures than would be the case for an ideal system. The vapor pressure curves are of the form shown in Fig. 161; it will be observed that the total vapor pressure may well have a minimum value for a particular mixture. Examples of systems of this type are: pyridine and formic, acetic or propionic acids, i.e., a basic and an acidic constituent; and mixtures of a halomethane, e.g., chloroform, with an oxygen or nitrogen compound, e.g., a ketone, ether, ester or amine. In every case there are reasons for believing that there is partial

compound formation in the liquid, the two molecules being connected through a hydrogen bond. Negative deviations from Raoult's law also arise with mixtures of a strong volatile acid, e.g., the halogen acids and nitric and perchloric acids, with water. It is probable that in these systems hydrates are formed in solution, although interaction leading to the formation of nonvolatile ions in place of the volatile acid molecules is mainly responsible for the negative deviations.

As with solutions which exhibit positive deviations, the experimental partial vapor pressure curves approach the ideal ones as the system approaches pure A or pure B. At high dilution, therefore, the "solvent" obeys Raoult's law, while the "solute" satisfies Henry's law.

Effect of Temperature.—Increase of temperature almost invariably has the effect of making any system approach more closely to ideal be-

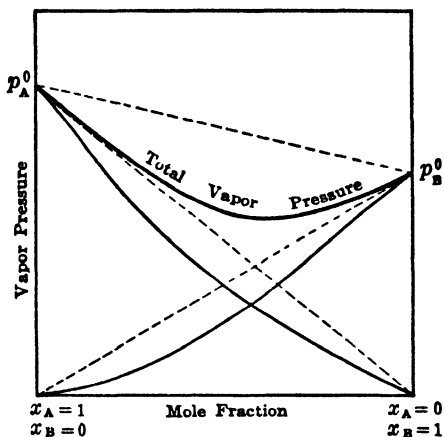


FIG. 161. Negative deviations from Raoult's law

havior, whereas diminution of temperature increases the deviations; changes in the form of the vapor pressure curves are, therefore, to be expected as the temperature is altered. Thus, a system which shows small positive deviations from Raoult's law will have a total vapor pressure curve which is concave downwards but, as the temperature is lowered, this may become one with a maximum. Eventually, at still lower temperatures, the maximum may become broad and flat, and finally the liquid may separate into two layers.

Composition of Liquid and Vapor.—From the vapor pressure curves already described, the composition of the vapor in equilibrium with any particular liquid at a given temperature can be found, making the assumption that the mole fraction of each constituent in the vapor is proportional to its partial vapor pressure (see p. 303). Important general conclusions can, however, be reached from purely theoretical considerations. The Duhem-Margules equation may be written in the form

$$\frac{x_A}{p_A} \cdot \frac{dp_A}{dx_A} = \frac{x_B}{p_B} \cdot \frac{dp_B}{dx_B} \quad (46)$$

and since $dx_A = -dx_B$, this becomes

$$\frac{x_A}{p_A} \cdot \frac{dp_A}{dx_A} + \frac{x_B}{p_B} \cdot \frac{dp_B}{dx_A} = 0, \quad (47)$$

$$\therefore \frac{dP}{dx_A} = \frac{dp_A}{dx_A} + \frac{dp_B}{dx_A} = \frac{dp_B}{dx_A} \left(1 - \frac{x_B p_A}{x_A p_B} \right), \quad (48)$$

where P is the total vapor pressure. The value of dp_B/dx_A , which is equal to $-dp_B/dx_B$, must be negative; hence if dP/dx_A is to be positive, it follows that

$$x_B p_A > x_A p_B \quad \text{or} \quad \frac{p_A}{p_B} > \frac{x_A}{x_B}, \quad (49)$$

and this means that the vapor is richer in A than is the liquid from which it vaporizes. Similarly, if dP/dx_A is negative, that is dP/dx_B is positive, the vapor will be relatively richer in B. The vapor is, therefore, relatively richer in the component whose addition to the liquid mixture results in an increase of total vapor pressure. This is a form of what is known as **Konowaloff's rule**; it was arrived at empirically by D. P. Konowaloff (1881), who was not aware that he had been anticipated by the theoretical deductions of J. W. Gibbs (1876). It is of interest to record that Konowaloff made the first systematic measurements of the total vapor pressures of homogeneous liquid systems, and showed that mixtures could be divided into three types, according as the vapor pressures either increased regularly with composition, or showed a maximum, or a minimum.

If the vapor pressure curve has a maximum or minimum, then dP/dx_A , in (48), must be zero; hence either dp_B/dx_A must be zero, or $x_A p_B = x_B p_A$,

i.e., $x_A/x_B = p_A/p_B$. The former condition is unlikely, since it would mean that the *partial* vapor pressure would remain constant in spite of a change of concentration, and so for a maximum or minimum total vapor pressure,

$$p_A/p_B = x_A/x_B. \quad (50)$$

The molecular ratio of the two components in the vapor is given by p_A/p_B , if the gas law is obeyed, and so it follows that when the vapor pressure is a maximum or minimum, the composition of the vapor is the same as that of the liquid with which it is in equilibrium.⁹

Distillation of Homogeneous Binary Mixtures: Boiling Points.—If distillation is carried out isothermally, then the composition of the distillate, which is the same as that of the vapor in equilibrium with the liquid, can be determined from the isothermal vapor pressure curves described above. Greater interest attaches, however, to distillation at constant pressure, generally 1 atm., the temperature varying as the composition of the boiling liquid changes. The boiling point of a mixture, when it distills freely, is the temperature at which the total vapor pressure is equal to that of the atmosphere, and according to the phase rule each mixture must have a definite boiling point. It can be readily seen that for ideal systems or for those departing only slightly from ideal behavior at the boiling point, giving vapor pressure curves without a maximum or minimum, the boiling point of the system will vary regularly in an analogous manner. If the curve has a maximum, then for a particular composition of the liquid there should be a minimum boiling point; similarly, when the system shows marked negative deviations, and the total vapor pressure curve passes through a minimum, there will be a mixture having a boiling point higher than that of any other. The boiling point-composition curves should thus approximate in type to the vapor pressure-composition curves inverted. These anticipations are borne out in practice, as may be shown by direct measurement of the boiling points of liquid mixtures under constant pressure. In connection with the problems of distillation, it is important to know not only the boiling point, but also the composition of the vapor in equilibrium with each particular liquid mixture.

Experimental Methods.—A variety of methods have been used of which some of the more important will be described. A simple procedure, although perhaps not very accurate, was adopted by T. Tyrer (1912). The liquid was boiled, its temperature was recorded, and at the same time some of the vapor was trapped in a bulb of known capacity. The bulb was then removed, cooled and weighed; from the temperature, pressure, volume and weight of the vapor its composition was determined on the assumption that the gas laws apply, the molecular weights of the constituents being known.

The apparatus described above (Fig. 158) for determining the composition of the vapor in equilibrium with boiling liquid, in which the vapor is condensed and an appreciable quantity collected, may be adapted to the present problem. Instead of adjusting the pressure so that the liquid boils at a predetermined

temperature, the pressure is kept at 1 atm., and the temperature raised until steady boiling occurs. When equilibrium is attained, the boiling point is recorded, and samples of boiling liquid and of condensate are withdrawn for analysis. The procedure is repeated with a series of mixtures of different composition. A modification of the apparatus, known as the Othmer still, devised by D. F. Othmer (1928, 1932) is particularly useful for this work.

If the mixed vapor of constant composition is passed through a liquid mixture of the same components, the composition of the liquid will change steadily until equilibrium is attained; if the pressure is 1 atm. the temperature reached will be the boiling point of the solution. This principle, which resembles that used in the Landsberger method for determining the boiling points of solutions of nonvolatile solutes (p. 639), was developed by M. A. Rosanoff (1909). The apparatus is similar to McCoy's modification of the Landsberger apparatus; the outer vessel contains a liquid whose composition is kept constant by gradual addition of small amounts of the more volatile constituent. On boiling, the liquid produces the vapor of constant composition which passes into the inner tube containing the liquid mixture under investigation. The temperature of the latter alters steadily until equilibrium is reached when it remains constant; this is the boiling point of the mixture in the inner tube. The vapor passing through is now condensed and sufficient collected for analysis by a suitable method; a sample of the liquid in the inner tube is also withdrawn and analyzed. Thus the composition of liquid and vapor in equilibrium at the boiling point can be obtained. By using a number of different mixtures in the inner and outer tubes the whole composition range can be covered.

It was discovered by F. D. Brown (1880), and verified by M. A. Rosanoff, *et al.* (1915), that when a liquid is distilled with a still-head maintained at constant temperature, the composition of the distillate is identical with that of the vapor evolved by a mixture whose boiling point is equal to the temperature of the still-head. By a series of measurements of the boiling points of liquids of different compositions, together with another series involving the analysis of the distillates obtained with the still-heads kept at different temperatures, all the information required for the present purposes can be obtained.¹⁰

Results.—The experimental data are best plotted on a diagram in which the compositions are abscissae and the boiling points are ordinates; two curves are then obtained, one giving the composition of the liquid and the other that of the vapor with which it is in equilibrium at the boiling point. The relation between the two curves is obtained by a rule which follows from that of Konowaloff (p. 714); the vapor phase is relatively richer in the component which results in a lowering of the boiling point when added to the mixture, or alternatively, the liquid phase is richer in the component which raises the boiling point. These rules have not only a theoretical basis, but are also in agreement with experiment. It will be interesting to consider the curves in three main groups corresponding to those into which Konowaloff divided the vapor pressure curves; they are (a) the boiling point rises steadily with change of composition from the more volatile to the less volatile constituent; (b) there is a minimum, and (c) there is a maximum, boiling point. The connection

between these three types and deviations from ideal behavior have been already mentioned.

(a) **The Boiling Point Increases Regularly.**—An example of the first type is shown by the curves in Fig. 162; in accordance with the rule given above, the upper curve represents the composition of the vapor phase and the lower that of the liquid, the vapor thus always contains more A, which has the lower boiling point. When a liquid of composition represented by l boils at temperature t , the composition of the vapor first evolved is given by v ; * since this is richer in A, the liquid remaining becomes relatively

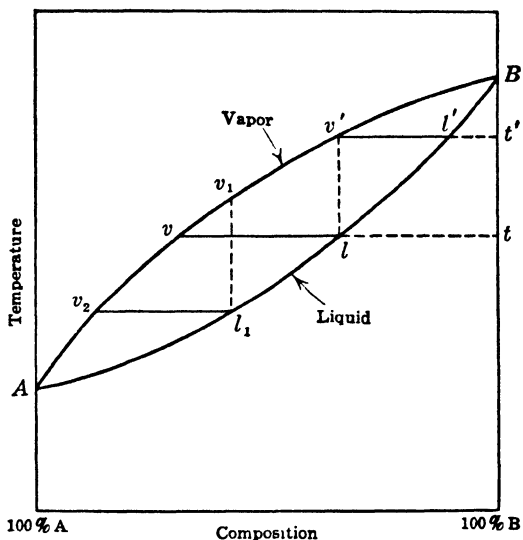


FIG. 162. System with regularly increasing boiling point

richer in B and its boiling point consequently rises.† This accounts for the gradual rise in the temperature observed during distillation. If the vapor were not allowed to escape, but kept in equilibrium at constant pressure with the liquid while boiling was continued, the composition of the liquid would move from l to l' and that of the vapor from v to v' , the boiling point rising from t to t' . At the latter temperature it is seen that the vapor in equilibrium has the composition of the original liquid; this means that at t' the whole of the liquid has been converted into vapor, l' representing the composition of the last drop of liquid.

* The line joining two points giving the compositions of systems in equilibrium is called a **tie line**.

† It is assumed in the present treatment that the vapor evolved is in equilibrium with the liquid; in normal ebullition this is not strictly true, since the formation of bubbles is due to superheating. The vapor consequently contains relatively more of the higher boiling point constituent than it should according to the equilibrium curve; this discrepancy, however, does not affect the main arguments.

In actual distillation this condition does not arise, for the vapor is continuously removed from the boiling liquid and condensed. Equilibrium is, therefore, never attained, and the composition of the liquid changes steadily from l to pure B, the boiling point rising from t to the temperature at which B boils. In fractional distillation the vapor passes up a suitable column and is then condensed in a series of fractions, as the boiling point of the liquid increases. The average composition of the first fraction will obviously be close to v , probably at v_1 if about five equal fractions are collected. If the first fraction is redistilled it will commence to boil at the temperature corresponding to l_1 , and the first vapor will have the composition v_2 , which again contains relatively more A, and once more the boiling point will rise. It is evident, however, that the distillate will be much richer in A than was the original liquid l , and if this process is repeated a few times a quantity of pure A will be obtained. The final fraction, or residue, of the first distillation will have a composition approaching that of B, and by redistilling this liquid an almost pure specimen of B can be obtained.

In practice additional advantage is gained by the use of a fractionating column, for as the vapor rises it meets a downflow of liquid which has a lower temperature; from Fig. 162 it will be evident that if the temperature of the vapor is lowered, it will partly condense, giving a liquid richer in B and leaving a vapor richer in A. The vapor passing up the column, therefore, contains even more A than did the vapor which left the boiling liquid; similarly, the liquid returning to the distilling flask will contain relatively more of B, the less volatile component. A mixture of two substances giving a boiling-point curve of the type under discussion can, therefore, be separated into its constituents by fractional distillation. As a corollary to Konowaloff's rule it may be stated that the greater the slope of the boiling-point curve the greater the difference in composition between liquid and vapor; hence, the greater the difference in the boiling points of the two constituents of a mixture the more easily can they be separated by distillation. Mixtures of this type are common, but the following may be mentioned: oxygen and nitrogen; carbon disulfide and benzene; benzene and toluene; acetone and water; acetone and ether; chloroform and carbon tetrachloride; and chlorobenzene and bromobenzene. All ideal systems belong to this category, but owing to the proximity of the boiling points separation of the components may not always be possible.

(b) **Minimum Boiling Point.**—Systems in which one mixture has a minimum boiling point give curves like those in Fig. 163; in accordance with Konowaloff's rule, the upper curve indicates the composition of the vapor phase. It will be seen that the two curves meet at the minimum M , so that here liquid and vapor in equilibrium have the same composition; this is to be expected theoretically from the Duhem-Margules equation (see p. 715) for a mixture having a maximum vapor pressure. The liquid represented by M will, therefore, boil at a constant tempera-

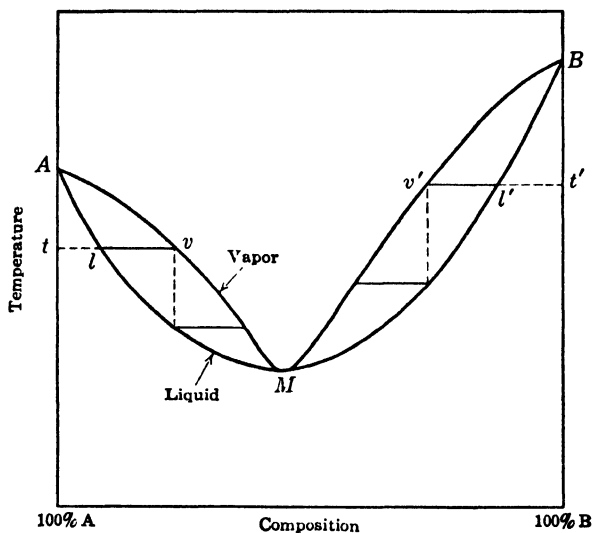


FIG. 163. System with minimum boiling point

ture and distil over completely without change of composition; such systems, which distil unchanged, have been called **azeotropic mixtures** (Greek: *to boil unchanged*) by J. Wade and R. W. Merriman (1911). The composition and boiling point of the constant boiling mixture vary with the pressure, as can be seen from the data in Table 122 on the

TABLE 122. COMPOSITIONS AND BOILING POINTS OF CONSTANT BOILING MIXTURES

Pressure	Ethyl Acetate-Water		Ethyl Acetate-Alcohol	
	B. Pt.	Water*	B. Pt.	Alcohol*
200 mm.	37.6° c.	5.79%	38.4° c.	20.5%
300	46.8°	6.56	47.8°	23.2
500	59.4°	7.54	60.6°	27.2
760	70.4°	8.43	71.8°	31.0
900	75.1°	8.80	76.6°	32.7
1500	90.3°	10.04	91.9°	39.1

* Compositions are in weight per cent.

systems ethyl acetate-water and ethyl acetate-ethyl alcohol. It may be noted that below 33° c., i.e., at pressures below 100 mm., the ethyl alcohol-water system does not have a minimum boiling point, although above this temperature and pressure it does.¹¹

The results to be expected when a system of the type having a boiling-point curve with a minimum is distilled may be seen from Fig. 163. When the liquid *l* boils at temperature *t* the vapor first evolved has the composition *v*, that is, it is richer in B; as distillation proceeds the composition of the liquid moves towards A and the boiling point rises. Provided the vapor is continuously removed, as it is in practice, the residue in the distilling flask tends towards pure A in composition. If the condensate is

redistilled, however, the vapor approaches the composition of the minimum boiling point system, as seen from the diagram. Fractional distillation will consequently result in a distillate of composition M , although the final residue will approach A ; separation into the two pure constituents is thus impossible. Similarly, if a liquid l' , which is to the right of the minimum, boils at t' the distillate will tend towards the composition of M , whereas the residue will approximate to B in composition. For systems having a minimum boiling point it is, therefore, only possible to obtain in a pure state by fractional distillation the component present in excess, which remains as the final residue; the distillate tends towards the mixture of minimum boiling point. The azeotropic mixture thus behaves in distillation as if it were a single substance, although the variation of its composition with pressure (Table 122) shows it to be a mixture. Examples of systems having boiling point curves with minima are very common; about four hundred are listed in the International Critical Tables. Some instances have already been mentioned and others are ethyl alcohol and benzene; methyl alcohol and chloroform; *n*-propyl alcohol and water; acetone and carbon tetrachloride; nitrous oxide and ethane (45 atm.); and many mixtures of a hydroxylic compound and a nonpolar substance. All these systems, as is to be expected, show marked positive deviations from the ideal behavior of Raoult's law, and give vapor pressure curves with maxima.

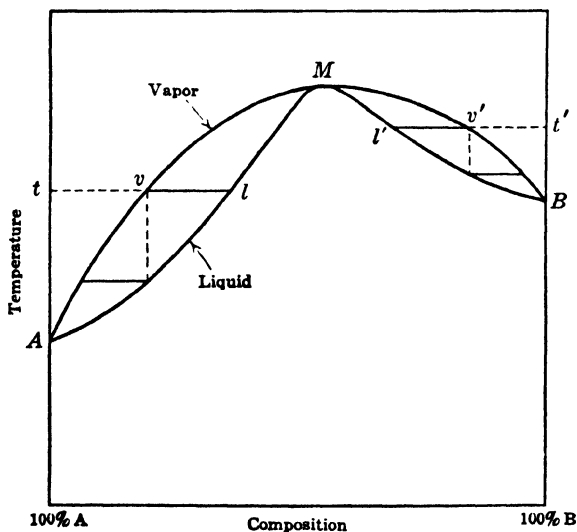


FIG. 164. System with maximum boiling point

(c) **Maximum Boiling Point.**—When the boiling point curve has a maximum, as in Fig. 164, the vapor-composition curve is still above that for the liquid phase. The two curves meet at the maximum point M

where again the liquid and vapor have the same composition; the liquid *M*, also called an azeotropic mixture, will therefore distil unchanged. Although systems with a maximum boiling point are not common, for they correspond to those having a vapor pressure curve with a minimum, i.e., showing negative deviations from Raoult's law, an azeotropic mixture of this type, viz., water and hydrogen chloride, was one of the first to be discovered (J. Dalton, 1832) and fully investigated. The boiling point of the mixture and the composition of the distillate are so constant that the system was at one time believed to consist of a compound of the two components; that this is incorrect is shown by the fact that the substances are not present in a simple molecular ratio, and also by the discovery of H. E. Roscoe and W. D. Dittmar (1860) that the composition of the constant boiling mixture varies with the pressure. At any given pressure, however, the composition of the distillate is so definite that, following the suggestion of G. A. Hulett and W. D. Bonner (1909), constant boiling hydrochloric acid is used for the purpose of making up standard solutions for volumetric analysis. The boiling points and compositions, in weight per cent, of the mixtures at various pressures are given in Table 123.

TABLE 123. CONSTANT BOILING MIXTURES OF HYDROCHLORIC ACID AND WATER

Press.	B. Pt.	HCl	Press.	B. Pt.	HCl
400 mm.	92.08° c.	21.235%	700 mm.	106.42° c.	20.360%
500	97.58°	20.916	760	108.58°	20.222
600	102.21°	20.638	800	110.01°	20.155

It is evident from Fig. 164, as in the previous case, that separation of the components of a mixture by fractional distillation will not be possible. When the liquid *l* boils at *t* the vapor *v* which first distils over is richer in A; the residual liquid becomes richer in B and its boiling point rises. Fractionation of the distillate will give ultimately, as can be seen from the figure, a specimen of pure A, but the residue will approach *M* in composition and will eventually distil unchanged. Similarly, if the liquid *l'*, boiling at *t'*, is distilled the vapor contains relatively more B so that this component can eventually be obtained in a pure form by re-distillation of the condensate. Again the residue in the distilling flask will tend to become the constant boiling mixture and distil without change of composition. It is thus possible to obtain in a pure state by distillation only the constituent present in excess of that required by the azeotropic mixture. All the cases mentioned on p. 713, as having vapor pressure curves with a minimum, give maximum boiling-point mixtures.¹²

Partially Miscible Liquids.—Not all liquids are completely miscible with each other; in between this extreme and that of almost complete immiscibility, e.g., mercury and water, there is an important type of system consisting of two liquids which are partially miscible. For example, if a little phenol, ether or aniline is added to water at ordinary temperatures, the material will dissolve completely, but if the addition is continued a point is reached when no further dissolution occurs and two

liquid layers form. One of these layers is a saturated solution of the organic substance in water, and the other is a saturated solution of water in the organic compound; the two systems in equilibrium are called **conjugate solutions**. It is not necessary for one of the substances to be water, as incompletely miscible systems are frequently obtained with two organic compounds, e.g., aniline and hexane, and carbon disulfide and methyl alcohol at ordinary temperatures, or with an organic compound and liquid sulfur or liquid phosphorus at higher temperatures. It is of interest to consider the conditions of equilibrium in such systems of partially miscible liquids, but before doing so an examination may be made of the factors which bring about their existence.

When two liquids differ in internal pressure, polarity, association, etc., the vapor pressures are greater than would be expected from Raoult's law, and the value of α in equations (39a and b), which is a measure of the departure from the law, is positive. As this quantity, which is constant for a given pair of liquids for a definite temperature, is increased, it can be easily shown that partial vapor pressure curves of the types I, II and III in Fig. 165 are obtained for each component. A set of curves of this

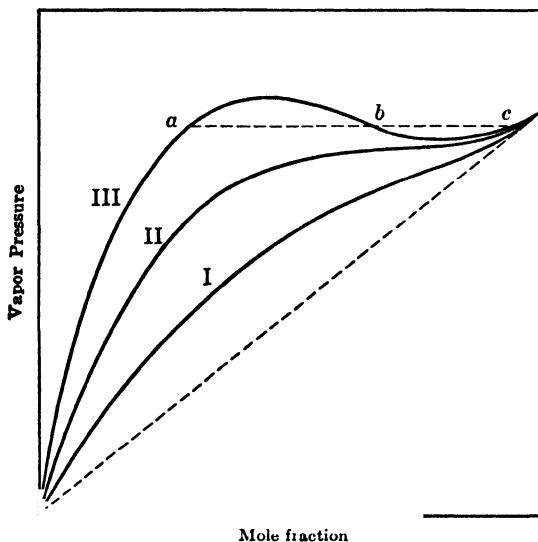


FIG. 165. Partial vapor pressures of system showing incomplete miscibility

nature results from a decrease in temperature of the system, or by changing one of the constituents, e.g., water and an alcohol, at constant temperature. Special interest attaches to curve III, for the \sim -shaped portion implies that there are three different solutions, *a*, *b* and *c*, having the same vapor pressure; in other words, three liquid layers should, apparently, be capable of coexisting in equilibrium. Application of the phase rule,

however, shows this to be impossible for a two component system. In order to define the state of the system it is necessary to fix, at least, the temperature; hence it follows from $F = C - P + 2$, that is

$$1 = 2 - P + 2 \quad \therefore P = 3,$$

that there cannot be more than three phases in equilibrium, and as one of these is vapor, two is the maximum number of liquid phases. The curve *abc* is evidently equivalent to the similar shaped portion of the van der Waals curves (p. 435), and in practice it is found to be flat, the liquid phase *b* having no real existence. The vapor pressure of the mixture is constant between *a* and *c*, and so temperature is the only variable; the system therefore consists of vapor and two liquid phases. Large positive deviations from ideal behavior thus lead to incomplete solubility of one liquid in another; in other words, a **miscibility gap** occurs. In the light of this conclusion, the partial miscibility of the mixtures recorded above is not unexpected. As already seen, temperature influences the nature of the partial vapor pressure curves, and hence the mutual solubilities of two liquids will be affected. Although some observations in this connection were made by D. Abasheff (1857), the important work in the field began with the investigations of W. Alexejeff (1875-86), and of F. Guthrie (1884).

Experimental Methods.—In the study of mutual solubilities of liquids it is generally the practice not to work at the vapor pressure of the system, but at the atmospheric pressure, and to determine the compositions of the two layers at various temperatures. For a system of two components in which there are two liquid phases in equilibrium, the vapor being neglected since it is not in equilibrium (cf. p. 744), the phase rule gives

$$F = C - P + 2, \\ \therefore F = 2 - 2 + 2 = 2.$$

The system is bivariant, and so if the pressure is fixed at 1 atm., the temperature alone will define the system completely. This means that at a given temperature the compositions of the two liquid layers are fixed, and are independent of the amounts of the two phases.

Three general methods have been employed in this work. In the analytical method the two liquids are brought to equilibrium by shaking together at a constant temperature; the two layers are allowed to separate and then portions are removed and analyzed. The measurements are repeated at a series of temperatures. As there is sometimes difficulty in finding suitable analytical procedures, the so-called "synthetic method" devised by Alexejeff is most frequently employed. Definite amounts of the components are weighed out and sealed into a small glass tube; this is then gently shaken and its temperature raised gradually. The temperature is then observed at which the two layers become one, or alternatively that at which one layer becomes two on cooling; the correct points are easily detected by the disappearance or appearance, respectively, of turbidity. For each mixture made up one mean temperature

is recorded, and provided the whole range of mixtures has been studied a complete solubility curve, to be described below, is obtained. As considerable pressures may be generated in the sealed tube, and the solubility may thus be affected, the observations are sometimes made with open tubes. An ingenious method, described by A. E. Hill (1923), which he called the "thermostatic method", is virtually a combination of analytical and synthetic procedures. Different amounts of the two liquids are weighed into each of two special graduated vessels, and equilibrium is attained by shaking at constant temperature; the liquids are allowed to settle out and their volumes are read. Suppose m_A and m'_A are the weights of the component A used in the two vessels, v_1 and v_2 are the volumes of the upper and lower layers in one vessel, and v'_1 and v'_2 the volumes in the other vessel. As seen above, the concentrations of the two layers, at constant temperature, are independent of the amounts of these layers; if the concentrations of the component A are c_1 and c_2 g. per unit volume in upper and lower layers, respectively; then

$$v_1c_1 + v_2c_2 = m_A \quad \text{and} \quad v'_1c_1 + v'_2c_2 = m'_A. \quad (51)$$

These equations allow the volume concentrations c_1 and c_2 to be calculated, and similar equations permit of the evaluation of the concentration of the component B in the two phases. Sufficient data are also available for the weight concentration to be obtained if required.¹³

Results.—If the compositions of the conjugate solutions, at each temperature, are plotted, most systems give a curve like that in Fig. 166,

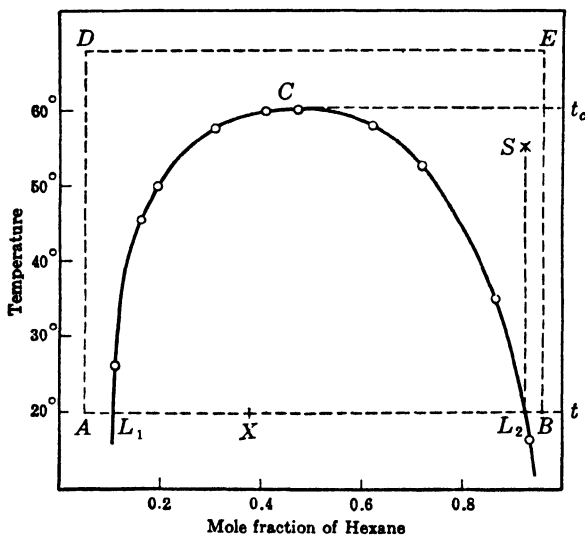


FIG. 166. The aniline-hexane system

which is actually that for an aniline-hexane mixture (D. B. Keyes and J. H. Hildebrand, 1917); at any temperature t the conjugate solutions have compositions indicated by the points L_1 and L_2 . The synthetic method,

described above, gives only one of the two points, representing the composition and temperature at which the second layer is negligible in amount; it gives, therefore, the composition of one of the two layers. Nevertheless by covering a sufficient range of concentration the whole curve may be mapped. The tie lines joining the points for the two layers in equilibrium at any temperature are, of course, horizontal.

It is seen from Fig. 166 that as the temperature is raised the compositions of the two layers approach each other; the mutual solubilities of the two components increase with increasing temperature indicating, as is generally the case, that the departure from ideal behavior is diminishing. At the temperature t_c the two layers have become identical in composition and are actually one layer, the composition of which is represented by C ; the temperature t_c is known either as the **critical solution temperature** (D. O. Masson, 1891) or the **consolute temperature** (W. D. Bancroft, 1894) of the system. Above this temperature, as first observed by Alexejeff, the two liquids become miscible in all proportions. In fact, all systems outside the curve consist of one liquid layer only, whereas those represented by points within the curve are made up of two layers, the compositions of which are given by points on the curve for the same temperature. Consider, for example, the system having a total composition indicated by the point X at temperature t ; this will consist of the two layers L_1 and L_2 , the relative amounts being given by

$$\frac{\text{Amount of } L_1}{\text{Amount of } L_2} = \frac{\text{Distance } XL_2}{\text{Distance } XL_1}. \quad (52)$$

The closer the point X is to L_1 the larger will be the proportion of that layer, and similarly for L_2 .

If to a liquid represented by A in Fig. 166, that is, to an unsaturated solution of hexane in aniline, quantities of hexane are added, the temperature being constant at t , the composition of the system will move along the horizontal line AL_1 . At L_1 the solution will be saturated and the second layer L_2 will just begin to appear; further addition of hexane will increase the proportion of the L_2 phase as compared with L_1 , but the compositions of the two layers will be unchanged. If sufficient hexane is added, it will eventually dissolve all the aniline, the second layer will disappear and an unsaturated solution of aniline in hexane, e.g., at B , will be obtained. Although it is convenient to speak of a solution of one substance in the other, there is no fundamental difference between the two components; this may be seen in the following manner. If the temperature of the solution A is raised above the consolute temperature t_c (59.6°) to D , excess of hexane added, so as to bring its composition to that of E , and then the whole cooled down until the original temperature is reached at B , a solution of hexane in aniline has been converted into one of aniline in hexane, although the system has never been divided into two layers. The temperature-solubility curve also gives information concerning the behavior of mixtures on heating or cooling at constant composition. If the one-phase system S is cooled it will remain homogeneous until the vertical line, indicating a lowering of temperature, cuts the curve at L_2 ; the second layer L_1 will then commence to form, and its presence will become evident by a turbidity. This is the principle involved in the "synthetic method" for obtaining the solubility curve. If the system is cooled further, the compositions of the two layers L_2 and L_1

will move down the right- and left-hand branches of the curve, respectively. If a homogeneous liquid having the critical composition *C* is cooled down from above the consolute temperature, the phenomenon known as **critical opalescence** will be observed as that temperature is reached. It consists of a bluish haze which is believed to be due to scattering of light brought about by local variations of density within the liquid (M. Smoluchowski, 1908).

Although the solubility curves do not often appear to be symmetrical, V. Rothmund (1898) has suggested that a law similar to that of Cailletet and Mathias (p. 431) might be applicable. If instead of the densities the mean compositions by weight, which are related to the densities, of the two layers are plotted for a series of temperatures the points frequently fall on a straight line passing through the consolute point *C*. Some workers prefer to plot the mean composition expressed in mole fractions, but it is not certain that this always gives better results.

A number of consolute temperatures, for systems of the same type as aniline and hexane, are quoted in Table 124; the column headed "compo-

TABLE 124. CONSOLUTE TEMPERATURES AND COMPOSITIONS

System	Temperature	Composition	System	Temperature	Composition
Methyl alcohol-Cyclohexane	49.1° c.	29%	Carbon dioxide-		
Isopentane-Phenol	63.5°	51	<i>n</i> -Butyl alcohol *	16°	65%
Carbon disulfide-Methyl			Water-Phenol	65.9°	34
alcohol	40.5°	80	Water-Aniline *	167°	15
Benzene-Sulfur *	163°	35	Aluminum bromide-		
			Sodium bromide	232°	92

* Under pressure.

sition" gives the amount of the *first named* component in weight per cent at the consolute point.¹³

It is of interest to mention that the critical solution temperatures of various hydrocarbons with aniline, or the miscibility temperatures of equal volumes of the two liquids, referred to as the "aniline points," are frequently used for the characterization and correlation of the properties of hydrocarbons.

Application of the Phase Rule.—All systems represented by points outside the solubility curve (Fig. 166) are found to consist of one liquid layer only; this is in agreement with the phase rule. If, as before, the vapor is neglected, there is only one phase, viz., liquid, and since there are two components, the phase rule shows the system to be trivariant. Since the pressure is taken as 1 atm., two degrees of freedom, namely temperature and concentration, still remain; both these quantities must, therefore, be fixed in order to define completely a system of one liquid layer. When there are two layers, however, there is only one degree of freedom, apart from pressure; that is to say, if the temperature is given, for example, the compositions of the two liquid layers are at once fixed by the points, e.g., *L*₁ and *L*₂, on the solubility curve.

Lower Consolute Temperatures.—In the course of his work Alexejff observed that for certain mixtures, e.g., ethyl ether and water, the mutual solubilities of the two components *decreased* with increasing tem-

perature, and a few years later V. Rothmund (1898) was able to show that some binary systems, e.g., triethylamine and water, and γ -collidine and water, had a *lower* critical solution temperature. As the temperature is lowered the mutual solubilities increase, and *below* a certain critical temperature the two liquids become miscible in all proportions. Mixtures behaving in this manner are uncommon, and reference will be made to some of them below. The general behavior of systems possessing a lower consolute temperature can be understood by applying the methods already considered for systems of the opposite type, the only difference being that the effect of temperature is reversed; the solubility curves are similar to that of Fig. 166 inverted. In those cases for which the effect of increasing temperature has been studied over a sufficient range, it is found that the solubility does not decrease indefinitely as the temperature is raised, but after a certain point it commences to behave in a normal manner and increase. Provided the critical temperature * of neither of the components is exceeded, for the liquid could then not exist, it is to be expected that systems possessing a lower should also have an upper consolute temperature. The solubility curve would then be a closed curve, all points outside it representing one layer, and inside it two layers. The first case of this type to be established was the system nicotine-water (C. S. Hudson, 1904), the consolute temperatures being 208° c. (upper) and 60.8° c. (lower); above the former and below the latter the liquids are miscible in all proportions. This may be seen from the solubility curve in Fig. 167. Although not all systems with a lower consolute temperature have been shown experimentally to possess an upper one, because it has not been possible to attain the requisite temperature, it is probable that they are fundamentally the same as those having two consolute points; all such systems may, therefore, be considered together, and some instances are quoted in Table 125. In addition to the cases mentioned, a

TABLE 125. SYSTEMS WITH LOWER AND UPPER CONSOLUTE TEMPERATURES

System	Lower Temp.	Upper Temp.	System	Lower Temp.	Upper Temp.
Nicotine-Water	60.8° c.	208° c.	4-Methyl piperidine-Water	85° c.	189° c.
Methyl ethyl ketone-Water	-6°	133°	Diethylamine-Water	143°	—
1-Methyl piperidine-Water	48°	>250°	Triethylamine-Water	18.5°	—
2-Methyl piperidine-Water	79°	227°	β -Picoline-Water	49°	153°
			γ -Collidine-Water	6°	225°
			Glycerol- <i>m</i> -Toluidine	6.7°	120°

number of glycol mono-ethers have been found to give upper and lower critical solution temperatures with water (H. L. Cox, 1926-27). In view of the high temperatures of the upper consolute points most of the systems were studied under pressure.

* The term critical temperature here refers to the ordinary critical point of the liquid-vapor system for each component; that is, the maximum temperature at which liquefaction is possible (p. 429).

The increased solubility observed with decreasing temperature suggests at first sight an unexpected approach to ideal behavior as the temperature is lowered; this is, however, probably not the case. It is true that the system is closer to the ideal, but this is due to the introduction at lower temperatures of negative deviations from Raoult's law which compensate the positive deviations already present. The negative deviations are undoubtedly brought about by compound formation between the two components of the system, and the extent of such com-

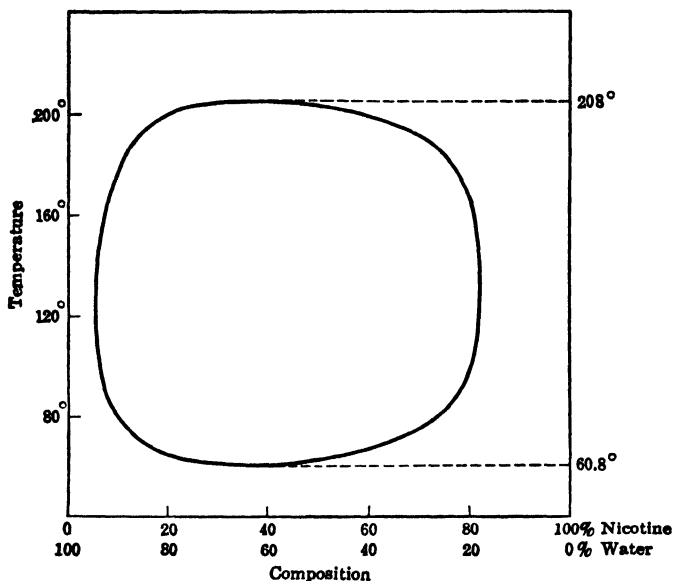
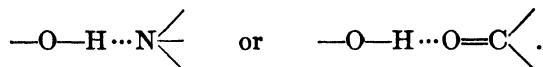


FIG. 167. The nicotine-water system

pound formation is expected to increase as the temperature is lowered. Systems possessing a lower consolute temperature almost invariably consist of either a hydroxy-compound and an amine, or a hydroxy-compound and a C=O (keto) or C—O—C (ether) grouping; compound formation appears to be connected with the presence of a hydrogen bond (p. 116); thus,



It is now generally accepted that even when a lower consolute temperature is not attained, the increase in solubility with decreasing temperature is an indication of compound formation between the two components of the system.¹⁴

Incomplete Systems.—Many pairs of liquids show neither upper nor lower consolute temperatures. The solubilities of ether and ethyl acetate, for example, in water increase with decreasing temperature, so that

a lower consolute point might be expected. The water freezes completely, however, before this temperature is reached, and so it is unattainable, at least at ordinary pressures. Similarly for the chloroform-water system an upper consolute point should be possible theoretically, but it would be above the critical temperature of chloroform; since the latter can then no longer exist as a liquid the system ceases to be one involving two liquid phases.

Influence of Added Substances.—The addition of a foreign substance to the system of two liquid components makes it become one of three components; the mutual solubilities of the liquids, like the solubility of a gas, depend on the nature and quantity of the added material. If the latter is soluble in only one of the two liquids, then it has been found that their mutual solubilities are decreased; in such instances the added substances will, in general, raise the consolute temperature. For example, a concentration of 0.1 mole of potassium chloride per liter of water raises the critical solution temperature of the phenol-water system by about 8° ; a similar concentration of naphthalene in the phenol increases the temperature by approximately 20° . It is of interest, in this connection, that the consolute point for this system was for many years believed to be 68.4°C. , whereas recent work has shown it to be about 66°C. ; the higher value has been attributed to the use of phenol containing a hydrocarbon impurity. Further, traces of water, which are soluble only in the polar component, produce a marked increase in the consolute temperature of systems involving a polar, e.g., alcohol, or acid, and a non-polar, e.g., hydrocarbon or carbon disulfide, liquid. Determination of the critical solution temperature may consequently be used as a test of purity (L. Crismer, 1895–96, 1904–06; D. C. Jones, 1923). When the added substance dissolves in both liquids the consolute temperature is lowered, as is the case when succinic acid is added to the phenol-water system. Other illustrations of the same type of behavior are the addition of alcohol to ethyl acetate-water and ether-water mixtures, which makes the components completely miscible, and the use of soap to increase the mutual solubilities of phenols and cresols in water in the preparation of commercial disinfectant solutions.

Salts are usually insoluble in organic liquids, and so the consolute temperature of systems consisting of water and an organic substance is generally raised by the presence of a salt; this is an illustration of the salting-out effect, to which reference has been made earlier (p. 699). An extreme example of the phenomenon is the formation of two liquid layers when a very soluble salt, such as potassium carbonate, is added to a mixture of ethyl alcohol and water. The influence of electrolytes on the solubility of a nonelectrolyte, e.g., an organic liquid, in water is exactly analogous to their behavior with gases, and both phenomena are probably to be ascribed to the same fundamental cause; an equation of the type of (9) is also applicable in many instances. Hydration numbers, which may be regarded at least as a measure of the salting-out effect, of various

ions have been evaluated and the results found to be in agreement with those obtained by measurements of the solubilities of gases in salt solutions. Some salts, particularly the lithium halides, are soluble in many organic liquids as well as in water; these should have a negative salting-out influence, and lower the consolute temperature. It has been found, in fact, that aniline and water become completely miscible in all proportions at ordinary temperatures if sufficient lithium iodide is added (S. Glasstone, 1927), although the consolute point in the absence of the salt is at 167°C .

The critical solution temperature of a system is changed if the isotopic ratio of the hydrogen, e.g., in water, is changed; for phenol and water, for example, the consolute temperature is raised 12.3° if deuterium oxide, i.e., heavy water, is used instead of ordinary water which consists almost entirely of the lighter isotopic form. The determination of the consolute temperature with phenol has been proposed as a method for estimating the proportions of the two hydrogen isotopes in water.¹⁵

Vapor Pressure Curves.—It has been seen that the partial vapor pressure curve for each component of a two liquid layer system has a discontinuity with a flat portion (p. 722); the total pressure curves are, therefore, similar in shape. In agreement with theoretical anticipations three types have been observed, all of which are the result of positive deviations from ideal behavior; * these are shown in Fig. 168 for a con-

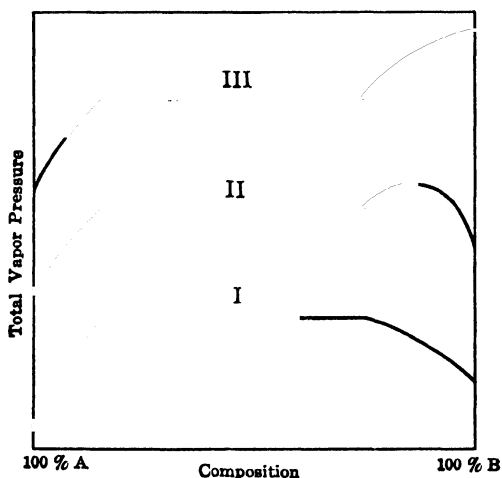


FIG. 168. Vapor pressure curves for partially miscible systems

stant temperature. (The relative positions of the curves in the figure have no significance; they are arranged in the particular order for convenience.) Systems giving curves of type I are aniline and water, isobutyl

* Negative deviations would not be expected to lead to two layers, unless some disturbing factor, such as association, were involved, e.g., in the triethylamine-acetic acid system.

alcohol and water, and ethyl acetate and water; examples of type II are water and phenol, methyl ethyl ketone and water, and triethylamine and water; curves of type III are given by isopentane and nitrobenzene, isopentane and dichloroacetic acid, and nicotine and water. The temperatures are, of course, within the range in which two liquid layers are formed; above the consolute points, when the two components are completely miscible, it is clear that systems I and II will give continuous curves with maxima as in Fig. 160.

The curves in Fig. 168 all consist of three parts: two in which the vapor pressure changes with the composition and the other in which it is independent of the relative amounts of the two constituents. The former are given by unsaturated solutions, that is, when there is only one liquid layer, and the latter result when two liquid phases are present. This can be readily proved by means of the phase rule; if there is one liquid and one vapor phase P is 2, and hence, since there are two components,

$$F = C - P + 2 = 2 - 2 + 2 = 2.$$

The system is bivariant and, although the temperature is fixed, the vapor pressure will still depend on one variable, namely, the composition of the liquid. When the two liquid layers are present, P becomes three and consequently F is reduced to one, and the system is univariant; hence, if the temperature is given the system is completely defined. In other words, as long as two liquid layers are present, the vapor pressure will be constant at a definite temperature irrespective of the relative amounts of the two layers. This is to be expected, of course, since the composition of each layer does not vary, although that of the mixture as a whole may change.

It is somewhat surprising, at first sight, that the partial pressures of each component are the same in two liquid layers at equilibrium. For example, consider a mixture of chloroform and water which separates into two layers; then the partial pressure of chloroform above the lower layer containing over 99 per cent is the same as that above the upper layer containing less than 1 per cent of chloroform. This result, which was verified experimentally by D. P. Konowaloff (1881), and can be shown to be a theoretical necessity, emphasizes the considerable deviations from Raoult's law shown by a system of partially miscible liquids. Imagine an annular vessel containing the two liquids A and B in equilibrium, as shown in Fig. 169; it is obvious that unless the vapor pressure of A, for example, were the same over both layers this component would distil from one to the other. The same argument applies to the vapor pressure of the component B; hence the partial pressure of each constituent must be the

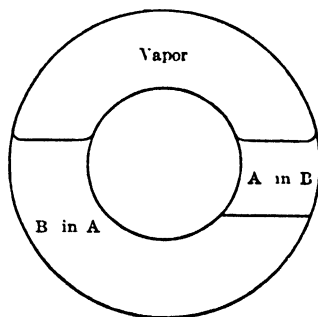


FIG. 169. Equilibrium between two liquid layers

same in the two layers. The same conclusion may be readily reached by a consideration of the chemical potentials of the components, remembering that the value must be the same for a given component in both layers.

Distillation.—Few complete results appear to have been published of the boiling points and compositions of liquid and vapor for systems of the type under discussion. It is, however, not difficult to infer from general considerations, and from Konowaloff's rule, as well as from some experimental evidence, that three types of curves, corresponding to the vapor pressure curves, must exist; these are shown in Fig. 170 numbered to cor-

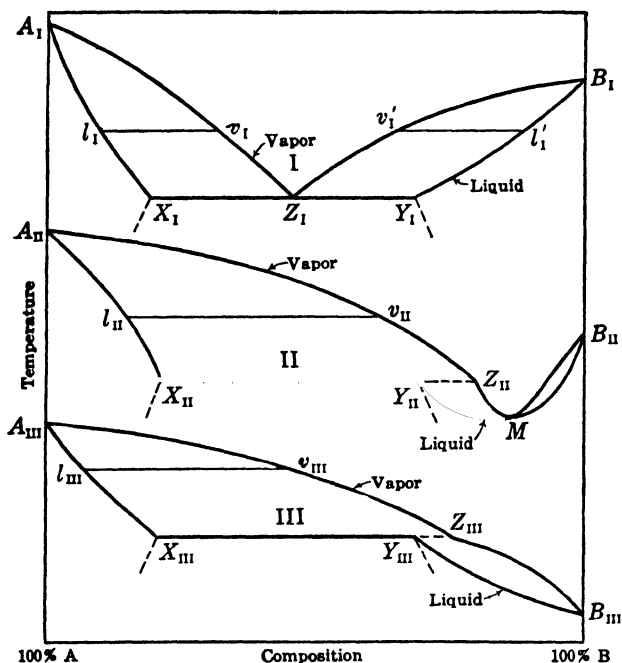


FIG. 170. Distillation of partially miscible liquids

respond to the curves in Fig. 168. In each case the upper curve gives the composition of the vapor in equilibrium with the liquid at the boiling point. When the boiling point varies with composition the system consists of one liquid phase; the horizontal portion, i.e., constant boiling point, represents two liquid phases. For the latter there is only one degree of freedom, as shown above, and so if the pressure is fixed, e.g., 1 atm., the boiling point will be independent of the total composition.

From the argument on p. 717 it will be clear that when a liquid l_1 in Fig. 170, I is distilled the vapor will be richer in component B; the boiling point will

therefore rise and the residue will eventually approximate to A. By suitable fractionation the latter could be obtained almost pure. If the vapor v_I is condensed and fractionally distilled, the volatile fraction will eventually have the composition Z_I , and will separate into two layers X_I and Y_I . This system will distil over unchanged, since at Z_I the total composition of the liquid is the same as that of the vapor with which it is in equilibrium. Fractionation of l'_I would, similarly, give pure B as the residue and the system Z_I as the most volatile fraction. If any two-liquid system, that is of total composition lying between X_I and Y_I , is distilled, the boiling point will be constant and the vapor composition will be given by Z_I . If the original system contained more A than does Z_I , that is its composition was between X_I and Z_I , then in the course of distillation the layer Y_I will disappear; after this the boiling point of the remaining layer X_I will rise to that of pure A, the compositions of the liquid and distillate changing correspondingly. Similarly, if the total composition is initially between Z_I and Y_I , the first distillate will have the composition Z_I ; eventually the layer X_I will disappear, and the boiling point will rise slowly to that of pure B.

In case II the liquid l_{II} will give the vapor v_{II} , and the boiling point will rise to that of pure A; continued fractionation of the distillate will result eventually in the most volatile fraction of composition M , the mixture with the lowest boiling point, being obtained. Further fractionation is then impossible for the system distills unchanged, as shown by the curves for liquid and vapor touching at this point. If any mixture between X_{II} and Y_{II} , that is consisting of two liquid layers, is distilled it will boil at constant temperature and give vapor Z_{II} ; since this contains relatively more B than does either layer, the composition of the remaining liquid moves to the left. The Y_{II} layer will thus eventually disappear leaving X_{II} , the boiling point and composition of which will then rise to A. The boiling points of liquids between M and Y_{II} will rise on distillation and the composition of the residue will move towards Y_{II} ; at this point two layers X_{II} and Y_{II} will be formed and the distillation will continue as already described for such systems. The behavior of systems between M and B_{II} is identical with that for a system of two completely miscible liquids having a mixture of minimum boiling point (p. 718). It will be noted that no matter what mixture of A and B is fractionated the composition of the vapor will tend towards that of the azeotropic mixture M of minimum boiling point.

If a liquid l_{III} , as in Fig. 170, III, is distilled, the boiling point will rise to that of pure A, and fractionation of the distillate will lead eventually to the most volatile fraction being pure B. Separation of the two constituents of any mixture by fractional distillation is here possible, in spite of the formation of two layers. Distillation of a two-layer system, that is between X_{III} and Y_{III} , will give vapor of composition Z_{III} ; here again the layer Y_{III} must eventually disappear leaving X_{III} , and the boiling point and composition of the liquid will then rise to that of pure A. The behavior of mixtures between B and Y_{III} will be similar to that between Y_{II} and M in case II; the final residue will be pure A whereas fractionation of the distillate will give B as the most volatile fraction.¹⁶

Completely Immiscible Liquids.—The total vapor pressure of a system of two almost completely immiscible liquids, such as carbon disulfide and water, is the sum of the separate vapor pressures of the two pure com-

ponents of the mixture. This was proved experimentally by H. V. Regnault (1854) who obtained the following results (Table 126) for the

TABLE 126. VAPOR PRESSURES OF CARBON DISULFIDE-WATER MIXTURES

Temp.	Pure H ₂ O	Pure CS ₂	Sum	Mixture
12.07° c.	10.5 mm.	216.7 mm.	227.2 mm.	225.9 mm.
26.87°	26.3	388.7	415.0	412.3

carbon disulfide-water system; the pressures are in mm. of mercury. The observed value for the mixture is somewhat less than the sum of the individual pressures, as might be expected if the liquids were slightly soluble in each other. The total pressure of the system is, of course, independent of the relative amounts of the two liquids; each exerts its own vapor pressure regardless of the presence of the other. The composition of the vapor can be easily calculated on the assumption that the gas laws are obeyed, for then the number of molecules of each constituent in the vapor will be proportional to its partial pressure, i.e., to the vapor pressure of the pure liquid at that temperature. If p_A and p_B are the vapor pressures of the liquids A and B in the pure state, and n_A and n_B the number of moles of each in the vapor, then if the liquids are completely immiscible, the total pressure P is given by

$$P = p_A^0 + p_B^0, \quad (53)$$

and the composition of the vapor by

$$n_A/n_B = p_A^0/p_B^0. \quad (54)$$

To express the ratio of A to B in the vapor in terms of the actual weights w_A and w_B , the quantities n_A and n_B must be multiplied by the respective molecular weights, M_A and M_B ; hence,

$$\frac{w_A}{w_B} = \frac{M_A n_A}{M_B n_B} = \frac{M_A p_A^0}{M_B p_B^0}. \quad (55)$$

A system of the type under discussion will boil when the total vapor pressure P is equal to that of the atmosphere; the boiling point of the mixture will thus be lower than that of either constituent because its vapor pressure is higher at all temperatures. Further, since the total vapor pressure is independent of the relative amounts of the two liquids, the boiling point will remain constant as long as the two layers are present; this, of course, can be readily deduced from the phase rule, as on p. 731.

The properties described above are utilized in the process of **steam distillation**, whereby a substance virtually immiscible with water, and having a relatively high boiling point, can be distilled at a much lower temperature by passing steam through it. The same result should, theoretically, be obtained by boiling a mixture of water and the particular immiscible substance, but by bubbling steam through the latter the system is kept agitated and equilibrium is attained between the vapor and

the two liquids. The mixture boils, that is it distils freely, when the total pressure of the two components is equal to that of the atmosphere. For example, when the external pressure was 740.2 mm. a mixture of water and chlorobenzene distilled at a constant temperature of $90.3^\circ \pm 0.05^\circ \text{C}$.; at this temperature the vapor pressures of the two constituents are 530.1 and 210.1 mm., respectively, making a total of 740.2 mm. The composition of the distillate will be equal to that of the vapor; hence,

$$\frac{\text{Weight of chlorobenzene } (w_A)}{\text{Weight of water } (w_B)} = \frac{M_A p_A^0}{M_B p_B^0} = \frac{112.6 \times 210.1}{18 \times 530.1} = 2.48.$$

The distillate should, therefore, contain 71.2 per cent of chlorobenzene by weight; the actual value found by experiment was 71.4 per cent. The agreement is remarkably good in view of the assumption involved that the vapors obey the gas laws. Chlorobenzene has a boiling point of 132°C ., yet when distilled with steam at a temperature 42° lower, the distillate contains over 70 per cent of the organic compound. An examination of (55) shows that the high proportion is due to the high molecular weight of chlorobenzene as compared with that of water; this case is a particularly favorable one in addition, because of the relatively high vapor pressure of the former in the region of 90° to 100°C . Aniline is generally distilled in steam in practice; under atmospheric pressure the system should boil at 98.5°C ., when the vapor pressures of aniline and water are 43 mm. and 717 mm. respectively. Taking the molecular weights as 93 and 18, it is calculated that the distillate should contain about 23 per cent of aniline by weight; the actual proportion is somewhat lower because aniline is appreciably soluble in water.

DISTRIBUTION IN LIQUID-LIQUID SYSTEMS

The Distribution Law.—If to a system of two liquid layers, made up of two immiscible or slightly miscible components, is added a quantity of a third substance soluble in both layers, then the substance is found to distribute, or divide, itself between the two layers in a definite manner (M. Berthelot, 1872). Suppose c_I and c_{II} are the concentrations in the layers I and II respectively, then at constant temperature

$$c_I/c_{II} = \text{constant } (K), \quad (56)$$

independent of the total amount of dissolved substance present; the latter distributes itself between the two layers in a constant ratio. This conclusion constitutes the **distribution law** or **partition law**, and the constant K is referred to as the distribution, or partition, coefficient. The law is of wide applicability and, as mentioned on p. 697, Henry's law for the influence of pressure on the solubility of a gas is a form of the distribution law; the gas may be imagined as distributing itself between free space, i.e., a vacuum, and the solvent in a constant ratio.

An interesting result is obtained by supposing the two liquid layers to be in contact with solid solute, so that when equilibrium is obtained both solutions are saturated; if the saturation solubilities are s_I and s_{II} in the two layers, then the distribution law gives

$$c_I/c_{II} = s_I/s_{II} = K, \quad (57)$$

so that the distribution coefficient is equal to the ratio of the saturation solubilities in the two layers. It will be seen below that the distribution law is theoretically only applicable to dilute solutions, and so this extension of the law must be regarded as in the nature of an approximation.

Application of the phase rule to the system of three components shows that since there are only two phases, the vapor not being in equilibrium, the system has three degrees of freedom; one of these must be the pressure of the atmosphere, and so at a definite temperature, which fixes another, only one variable determines the composition of the whole system. If the concentration of the solute in one layer is known, therefore, that in the other must be fixed, and this is in agreement with the distribution law. The law can, however, be deduced simply and more explicitly, by means of the rule that the chemical potential of any substance must be the same in two phases in equilibrium (p. 477). In layer I the chemical potential of the solute may be written

$$\mu_I = \mu_I^0 + RT \ln a_I, \quad (58)$$

and in layer II it is

$$\mu_{II} = \mu_{II}^0 + RT \ln a_{II}, \quad (59)$$

where a_I and a_{II} are the activities in the two layers. At equilibrium the two values of the potential are equal, and since μ_I^0 and μ_{II}^0 are constants, at definite temperature and pressure, it follows that

$$a_I/a_{II} = \text{constant}. \quad (60)$$

This is the exact expression of the distribution law, but for practical purposes it is put into an approximate form. If the solutions behave ideally, in the sense that either Henry's law or Raoult's law applies, the activities in (60) may be replaced by the respective mole fractions, i.e.,

$$x_I/x_{II} = \text{constant}.$$

Further, if the solutions are dilute the ratio of the mole fractions is almost the same as that of the concentrations, either in moles per liter of solution or in moles per 1000 g. of solvent; so that

$$c_I/c_{II} = \text{constant } (K) \quad (61)$$

at a definite temperature. This derivation of the distribution law shows that in its original form [equation (56)] it is only an approximation, based on the assumptions of dilute solutions and ideal behavior; another assumption involved is that the solute does not affect the miscibility of the two layers.

In spite of the theoretical limitations many substances have been found to obey the distribution law in a reasonably satisfactory manner,

provided the solute has the same molecular weight in both solvents. Some examples of the applicability of the law are given in Table 127: the concentrations are in moles per liter.

TABLE 127. DISTRIBUTION COEFFICIENTS

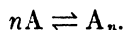
I ₂ between CCl ₄ (I) and H ₂ O(II)		SO ₂ between H ₂ O(I) and CHCl ₃ (II)		HgCl ₂ between H ₂ O(I) and C ₆ H ₆ (II)	
<i>c</i> _I	<i>c</i> _I / <i>c</i> _{II}	<i>c</i> _I	<i>c</i> _I / <i>c</i> _{II}	<i>c</i> _I	<i>c</i> _I / <i>c</i> _{II}
0.02	85.1	0.1	1.20	0.001	11.97
0.04	85.2	0.2	1.06	0.005	12.31
0.06	85.4	0.4	0.98	0.01	12.73
0.08	86.0	0.7	0.94	0.015	13.16
0.10	87.5	1.1	0.92	0.025	14.01

Association and Dissociation of Solute.—The emphasis on the necessity for the solute to have the same molecular weight in both layers is due to W. Nernst * (1891); he suggested that if in one or both liquids the solute consists to some extent of molecules of different weight or composition, then the distribution law is applicable only to the molecular species common to both phases. Consider, for example, the distribution of benzoic acid between water and benzene; in the former liquid the acid is ionized and consists of C₆H₅COOH molecules and C₆H₅COO[−] and H⁺ ions, while in benzene it is associated and is in the form of C₆H₅COOH and (C₆H₅COOH)₂ molecules. According to Nernst, the distribution law applies only to each individual species; if these are taken as C₆H₅COOH, then the concentration terms do not refer to the total concentration, but to these particular molecules. In general, if α represents the fraction of the total amount of solute which is dissociated or associated, assuming that either one or the other occurs in each solvent, then

$$\frac{(1 - \alpha_I)c_I}{(1 - \alpha_{II})c_{II}} = \text{constant}, \quad (62)$$

since $1 - \alpha$ gives the fraction of unchanged molecules in each case. This equation cannot be expected to hold very exactly, since solutions in which dissociation or association occurs depart from ideal behavior (p. 678).

Suppose a given solute A has a normal molecular weight in a solvent I but in the other solvent II it associates to give the complex molecule A_n; in the latter solvent an equilibrium will be set up between simple and associated molecules; thus,



If c_{II} is the *total* concentration of A, and α_a represents the fraction of the total which has undergone association, then $(1 - \alpha_a)c_{II}$ is the concentra-

* A somewhat similar point of view was put forward independently by P. Aulich (1891).

tion of simple molecules and $\alpha_a c_{II}/n$ is that of complex molecules. Application of the law of equilibrium (p. 817) leads to the equation

$$k_a = \alpha_a c_{II}/n(1 - \alpha_a)^n c_{II}^n, \quad (63)$$

$$\therefore (1 - \alpha_a) c_{II} = \sqrt[n]{\alpha_a c_{II}/n k_a}, \quad (64)$$

where k_a may be called the association constant. Assuming the distribution law to apply to the simple molecules only, it follows that

$$\frac{c_I}{(1 - \alpha_a) c_{II}} = \frac{c_I}{\sqrt[n]{\alpha_a c_{II}/n k_a}} = K, \quad (65)$$

where c_I is the concentration in the solvent I and K is the distribution constant. If the molecules in solvent II are almost completely in the associated form, α_a will approximate to unity and (65) will become

$$\frac{c_I}{\sqrt[n]{c_{II}}} = \text{constant}, \quad (66)$$

since n , k_a and K are all constants. By means of this relationship it is possible to determine approximate values of n in some cases. For example, the following results were obtained for the distribution of benzoic acid between water (I) and benzene (II) at 6° C.; the concentrations are expressed in moles per liter.

c_I	0.00329	0.00579	0.00749	0.0114
c_{II}	0.0156	0.0495	0.0835	0.195
c_I/c_{II}	0.210	0.117	0.089	0.058
$c_I/\sqrt[n]{c_{II}}$	0.0263	0.0262	0.0259	0.0258

It will be seen that although c_I/c_{II} varies considerably, the quantity $c_I/\sqrt[n]{c_{II}}$ is approximately constant, showing that in benzene solution benzoic acid is almost completely in the form of double molecules. This argument, however, ignores the dissociation in the aqueous layer, and a more exact treatment is possible, as shown below.

If the solute dissociates in solvent I to form two ions, then application of the law of equilibrium (p. 954) gives

$$k_d = \frac{\alpha_d^2 c_I}{1 - \alpha_d}, \quad (67)$$

$$\therefore (1 - \alpha_d) c_I = \alpha_d^2 c_I^2 / k_d, \quad (68)$$

where k_d is the dissociation constant of the acid, c_I is its *total* concentration and α_d the degree of dissociation. The concentration of undissociated acid, to be used in the distribution law, is thus $(1 - \alpha_d) c_I$, where the value of α_d for any given concentration can be calculated from (67) provided k_d is known. Assuming the solute to be normal in solvent II, then the distribution law takes the form

$$\alpha_d c_I^2 / k_d c_{II} = K. \quad (69)$$

If, however, as is the case when benzoic acid is distributed between water and

benzene, the solute is dissociated in solvent I and associated to form double molecules, i.e., $n = 2$, in solvent II, then

$$\frac{(1 - \alpha_d)c_I}{(1 - \alpha_a)c_{II}} = \frac{\alpha_d^2 c_I^2 / k_d}{\sqrt{\alpha_a c_{II} / 2k_a}} = K, \quad (70)$$

where c_I and c_{II} are the total concentrations in aqueous and benzene layers, respectively. To test the applicability of (70), k_a and k_d should be known, and then α_a and α_d could be calculated for various values of c_I and c_{II} . Although k_d is available, k_a is uncertain, and an alternative method is adopted: (70) is assumed to be correct, and with its aid, together with the value of k_d , the association constant k_a is determined for a series of concentrations as shown below. If constant results are obtained the arguments may be regarded as verified; incidentally, from k_a the extent of association in different solutions can be calculated by (63). The evaluation of k_a is carried out as follows: by (70)

$$(1 - \alpha_a)c_{II} = (1 - \alpha_d)c_I / K, \quad (71)$$

$$\therefore \alpha_a c_{II} = c_{II} - (1 - \alpha_d)c_I / K. \quad (72)$$

Substituting these values in (63), putting $n = 2$, it follows that

$$k_a = [c_{II} - (1 - \alpha_d)c_I / K] / [2(1 - \alpha_d)^2 c_I^2 / K^2]. \quad (73)$$

If two different amounts of benzoic acid are distributed between water and benzene, two different values of c_I and two corresponding values of c_{II} are available; further, for each c_I the corresponding degree of dissociation α_d may be calculated, since k_d is known for benzoic acid, and so K may be evaluated. In two experiments at 10° c. W. S. Hendrixson (1897) obtained the following results:

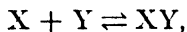
c_I	c_{II}	α_d	K
1.758 $\times 10^{-3}$ mole/l.	5.937 $\times 10^{-3}$ mole/l.	0.169	0.71
11.53	224.8	0.0695	

The dissociation constant of benzoic acid was taken as 6×10^{-5} , with concentrations in moles per liter. Assuming K to be 0.71, it is possible from (73) to evaluate k_a for these two cases, and also for a number of others in which different amounts of benzoic acid are employed. The results obtained are given in Table 128; the agreement between the various values for k_a may be regarded as supporting the views described above concerning the applicability of the distribution law.

TABLE 128. DISTRIBUTION OF BENZOIC ACID BETWEEN WATER (I) AND BENZENE (II)

c_I Moles/Liter	c_{II} Moles/Liter	α_d	k_a
2.304 $\times 10^{-3}$	9.76 $\times 10^{-3}$	0.149	459
4.608	36.24	0.104	450
7.299	89.28	0.0866	454
9.964	166.2	0.0746	455

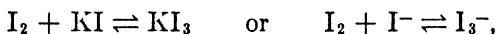
Study of Equilibria.—Distribution measurements have been used to study other aspects of equilibria in solution. If there is, in general, a reversible reaction between X and Y and their product XY in a given solvent, thus



then it is necessary to find another solvent, immiscible with the first, in which only one of the reacting substances, e.g., X, dissolves, and consequently in which the reaction cannot occur. Suppose the reaction takes place in the solvent II, but not in I, and the normal distribution ratio of X between the two solvents, in the absence of any other substance, that is c_I/c_{II} , is K ; this is assumed to remain unchanged, as far as the simple X molecules are concerned, when Y is added. Suppose the solvent II containing X and Y is shaken with solvent I until equilibrium is established; analysis then shows the total concentration of X to be c_I in layer I, and c_X in layer II, whereas that of Y is c_Y in layer II, none dissolving in I. If the distribution law holds, then the concentration of *free* X in solvent II must be c_I/K , and hence that in the form of XY will be $c_X - c_I/K$. The concentration of free Y in II will be the difference between the total c_Y and that as XY; that is $c_Y - c_X + c_I/K$. The equilibrium constant of the reaction is given by

$$\frac{\text{Concentration of XY}}{(\text{Conc. of free X})(\text{Conc. of free Y})} = \frac{c_X - c_I/K}{(c_I/K)(c_Y - c_X + c_I/K)} \quad (74)$$

An example of this type of application of the distribution coefficient is to the study of the reaction



which takes place in aqueous solution; the iodine is the only substance soluble in carbon tetrachloride or carbon disulfide, and one of these may be used as the second solvent. In the course of the study of this equilibrium, H. M. Dawson (1901), using carbon disulfide, obtained the following results at 13.5° c.

Total concentration of I_2 in aqueous layer = $c_X = 0.02832$ mole per liter

Total concentration of KI in aqueous layer = $c_Y = 0.125$ " " "

Concentration of I_2 in carbon disulfide layer = $c_I = 0.1896$ " " "

Distribution coefficient ($\text{CS}_2/\text{H}_2\text{O}$) = $K = 625$,

$\therefore c_I/K = 3.033 \times 10^{-4}$ mole per liter

$c_X - c_I/K = 28.02 \times 10^{-3}$

and

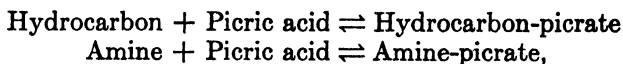
$c_Y - c_X + c_I/K = 96.98 \times 10^{-3}$ mole per liter,

$$\therefore \text{Equilibrium constant} = \frac{28.02 \times 10^{-3}}{(3.033 \times 10^{-4})(96.98 \times 10^{-3})} = 9.53 \times 10^2.$$

Similar values have been obtained in other experiments with different concentrations of iodine and potassium iodide. It should be noted that the distribution coefficient of iodine itself between the two solvents is obtained in separate experiments in which no potassium iodide is present.

The reaction between potassium bromide and bromine, to form perbromide, has been studied in an analogous manner. Investigations have

also been made of equilibria of the type



which occur in nonaqueous media, e.g., benzene or chloroform. Of the substances involved, only picric acid is appreciably soluble in water, and a study of its distribution between the organic solvent and water permits the equilibria to be studied. The reaction between chlorine and water to form hydrochloric and hypochlorous acids has also been studied, the concentration of free chlorine being determined by distribution with carbon tetrachloride (A. A. Jakowkin, 1896-99). Distribution measurements have been used to prove the existence of cuprammine ions, $\text{Cu}(\text{NH}_3)_4^{++}$, in an aqueous ammoniacal solution of copper sulfate, the distribution of the free ammonia being studied between chloroform and water (H. M. Dawson, 1906). The reactions between anhydro-bases, RNH_2 , and water, to form the hydrated bases RNH_2OH , have been investigated in a somewhat similar manner, the assumption being made that the latter are insoluble in chloroform (T. S. Moore, 1912). A distribution method has also been adapted to study the hydrolysis of salts, particularly of weak organic bases, since the free base is soluble in an organic solvent (see p. 990).

Distribution measurements have thrown light on the state of substances in different solvents. For example, if very concentrated aqueous hydrochloric acid solutions are shaken with benzene or nitrobenzene, appreciable amounts of hydrogen chloride pass into the nonaqueous layer. With less concentrated solutions, however, very little dissolves in the nonaqueous solvent. The result is attributed to the fact that only undissociated molecules are soluble in benzene or nitrobenzene; hence in the more dilute aqueous solution of the acid these can only be present to a minute extent and dissociation is almost complete (C. N. Hinshelwood, 1927; W. F. K. Wynne-Jones, 1930). When silver perchlorate is distributed between water and either benzene or toluene, the salt passes almost entirely into the water, indicating complete ionization in the aqueous layers (A. E. Hill, 1921); when aniline is used as the other solvent, however, the silver perchlorate is found almost entirely in this layer (A. E. Hill, 1924). Silver perchlorate and aniline form a number of compounds, and these must be so stable in aniline solutions that there is no species which can be common to it and water.¹⁷

Extraction With a Second Solvent.—The distribution of a solute between two partially miscible or immiscible liquids is the basis of the process of extraction with ether, or other solvent, frequently employed in organic chemistry. As already seen, the distribution coefficient is approximately equal to the ratio of the solubilities in the two solvents, and since organic compounds are usually relatively more soluble in organic solvents than in water, they may be extracted from aqueous solution.

If electrolytes are added to the water, the solubility of the organic compound may be depressed still further, that is, it will be salted-out, and the process of extraction will be facilitated.

It is possible to deduce a formula showing the efficiency to be expected from extraction under given conditions. Suppose v ml. of a solution (phase I) containing w g. of a dissolved substance are repeatedly extracted with fresh portions of l ml. of another solvent (phase II) which is quite immiscible with the first. If w_1 is the weight of solute remaining in phase I after the first extraction, then the concentrations are w_1/v g. per ml. in phase I and $(w - w_1)/l$ g. per ml. in phase II, so that the distribution coefficient K will be given by

$$\frac{w_1/v}{(w - w_1)/l} = K, \quad (75)$$

$$\therefore w_1 = w \frac{Kv}{Kv + l}. \quad (76)$$

After the second extraction w_2 g. remain in phase I, then the corresponding equation for the distribution coefficient gives

$$w_2 = w_1 \frac{Kv}{Kv + l} \quad (77)$$

$$= w \left(\frac{Kv}{Kv + l} \right)^2. \quad (78)$$

It is evident that this result can be generalized, so that after the n th extraction the weight w_n of solute remaining in phase I is

$$w_n = w \left(\frac{Kv}{Kv + l} \right)^n. \quad (79)$$

In practice it is desired to make w_n as small as possible for a given amount of extracting liquid, i.e., the product of n and l are constant; it is a simple matter to show that it is better to keep n large and l small, rather than the reverse. That is, the best results are obtained by a relatively large number of extractions with small amounts of liquid. It must be noted that the equations derived above assume immiscibility of the two liquids, so that they apply to extraction of an aqueous solution by a hydrocarbon or by chloroform; they will not hold exactly, however, for extraction by ether, but they nevertheless give a satisfactory qualitative indication of the results to be expected.

SOLID-LIQUID SYSTEMS

Freezing-Point and Solubility Curves.—The conditions of equilibrium between solid and liquid in a two-component system may be considered from two points of view. If the liquid mixture is in equilibrium with the solid phase of the component present in excess, that is, the substance

generally called the solvent, then the solution is said to be at its freezing point. The curve representing the variation of this temperature with the composition of the liquid phase is referred to as the **freezing-point curve**. On the other hand, if the solid phase of the substance present in relatively small proportion in the mixture is in equilibrium with the liquid, the latter is said to be a **saturated solution**, and the variation of composition with temperature is represented by a **solubility curve**. There is, however, no fundamental difference between solvent and solute, and the distinction between solubility and freezing curves is merely a matter of convenience. If both solids are appreciably soluble in the liquid mixture, however, then the two points of view become indistinguishable. For example, a liquid system containing 0.5 mole fraction of *p*-toluidine and 0.5 mole fraction of *o*-nitrophenol is in equilibrium with the latter solid at 17° c. It is equally reasonable to say that a solution of *p*-toluidine in an equimolar mixture with *o*-nitrophenol freezes at 17°, or that the solubility of *o*-nitrophenol in *p*-toluidine at this temperature is 0.5 mole fraction. There is no reason for preferring either statement over the other. The common practice is to use the description "freezing-point curves" for the conditions of equilibrium between solid and liquids of different compositions when the two components are similar chemically, e.g., when both are metals, salts or organic compounds of not very different melting point; it is then not possible to distinguish clearly between solvent and solute. If one of the constituents of the system differs from the other, e.g., salt and water, or if the two substances have different melting points, e.g., naphthalene and benzene, the one with the higher melting point never has a very high concentration in the liquid phase; it is then possible to make the conventional distinction between solvent and solute. The conditions under which the solid solute separates is indicated by a solubility curve, whereas the equilibrium between the liquid and the solid solvent is represented by a freezing-point curve.

Calculation of Freezing Points and Solubilities.—Strictly speaking the deduction of (34), p. 644, involves no assumption concerning the nature of the solid phase,* and so it applies equally to both components of the liquid mixture. Provided the latter behaves ideally, the general form of the equation is

$$\ln x = \frac{L_f}{R} \left(\frac{1}{T} - \frac{1}{T_0} \right) \quad (80)$$

for *constant pressure* conditions, where *T* is the temperature at which the solid present to the extent of *x* mole fraction in the mixture is in equilibrium with the liquid; *L_f* is the heat of fusion of the solid and *T₀* its melting point in the pure state. It is thus possible, by means of (80), to calculate the freezing point *T* of a solution in which *x* is the mole fraction of the solvent; alternatively, the equation gives the solubility, i.e., mole fraction *x*, at the temperature *T*. In each case *L_f* and *T₀* refer to the component separating as the solid phase.

* It must be a pure solid, and not a solid solution.

The approximate equation (37), p. 645, is applicable to the freezing points of dilute solutions, but for concentrated mixtures (80) gives better agreement with practice; deviations are observed, however, because the systems do not behave ideally. An important aspect of the equation is its application to the calculation of solubilities; this can be done provided the molar heat of fusion of the solute is known. For example, the heat of fusion of naphthalene is 4400 cal. per mole, and the melting point is 353° K., hence at 25° C., i.e., 298° K., it follows from (80) that x is 0.311. The saturation solubility of naphthalene in *any* solvent, *provided the solution behaves ideally*, should thus be 0.311 mole fraction at 25° C. In solvents such as benzene, toluene, chlorobenzene and chloroform, which have similar internal pressures to naphthalene, and are not specially polar, ideal mixtures with naphthalene might be anticipated (see Chapter IX); in these liquids the solubility is, in fact, very close to the ideal value. In aniline, acetone and hexane, whose internal pressures differ from that of naphthalene, the solubility is appreciably less than that calculated by (80); this corresponds to positive deviations from Raoult's law. Apart from this quantitative aspect a qualitative deduction of interest can be made: it is seen, according to (80), that for substances having similar molar heats of fusion, those with lower melting points have the higher solubilities.

Since T must always be less than the melting point T_0 , it follows from (80) that solubility must increase with temperature; further, the heat of solution must be equal to the heat of fusion of the solute, provided the system behaves ideally. By considering the equilibrium between a solid solute and its saturated solution, an equation of the form of (80) can be readily deduced for any system, ideal or not, L_f being replaced by the differential heat of solution (p. 241) in the saturated solution; for an ideal solution the two quantities, as seen above, are identical.¹⁸

Condensed Systems.—Since three degrees of freedom, viz., temperature, pressure and composition, are possible in a two component system, a complete graphical representation of the conditions of equilibrium would require the use of three dimensions. Solid models are frequently employed in this connection, but for most purposes a simplification is achieved by fixing arbitrarily one of the variables. With solid-liquid systems, which include from some points of view the most important aspects of phase equilibria, it is convenient to disregard the vapor phase and to fix an arbitrary constant pressure, e.g., 1 atm.; experimental studies are then made with vessels open to the atmosphere. It must be emphasized that since the pressure is generally not the equilibrium value, the system as a whole is not in true equilibrium, but for a solid-liquid system the effect of pressure is relatively small, and the measurements made under atmospheric pressure will be very little different from those which would be obtained if the solid and liquid had been in equilibrium under their own vapor pressure. Following the suggestion of J. H. van't Hoff (1894) a system in which only solid and liquid phases are considered

is called a **condensed system**. The graphical representation of conditions of equilibrium in a condensed system of two components is simplified by the fact that there are only two variables, temperature and composition, and these can be represented by ordinary rectangular coordinates. Solid-liquid equilibria of a number of types are known, and the more important will be described.

I. A. The Components are Completely Miscible in the Liquid State and the Solid Phases Consist of Pure Components.—If a liquid mixture of two components is cooled, solid will commence to separate at a definite temperature, namely, the freezing point. At this temperature there are two phases, viz., liquid and solid, the vapor being ignored, and since there are two components the system has two degrees of freedom; thus according to the phase rule

$$F = C - P + 2 = 2 - 2 + 2 = 2.$$

Of these one, the pressure, is arbitrarily fixed and so the condensed system is univariant; hence either the temperature or the composition of the liquid phase is alone sufficient to define the system completely. In other words, for every liquid mixture there will be a definite temperature, viz., the freezing point, at which it is in equilibrium with solid. If the freezing points of a series of liquid mixtures, varying in composition from one pure component A to the other B, are determined and plotted against the corresponding compositions of the liquid, two curves represented by *AC* and *BC* in Fig. 171 will be obtained. The points A and B are the freezing points of the pure components; the addition of B to A lowers the freezing point along *AC*, and similarly A added to B lowers its freezing point along *BC*, in accordance with the conclusion already reached (Chapter IX) concerning the influence of a dissolved substance on the freezing point of the solvent. If the two components A and B formed an ideal solution, the slope of the curves *AC* and *BC* could be calculated from the theoretical equations (p. 644) for the depression of the freezing points of concentrated solutions.

When liquids rich in A are cooled, that is between A and C, solid A separates, whereas liquids rich in B, between B and C, will deposit solid B on cooling. The curves *AC* and *BC* may thus be regarded as representing the conditions of temperature under which liquid mixtures of various compositions are in equilibrium with solid A or solid B, respectively. At the point C, where the two curves meet, obviously both solids A and B must be in equilibrium with the liquid; since three phases coexist the condensed system becomes invariant. There is thus only one temperature, at atmospheric pressure, where the liquid phase can be in equilibrium with both solids. The point C is clearly the lowest temperature at which *any* liquid mixture of A and B will freeze, and consequently represents the lowest melting point of *any* mixture of solid A and B; it has been called the **eutectic point** (Greek: *easily melting*) by F. Guthrie (1884), and is the lowest temperature for the existence of the liquid phase.¹⁹

Now that the significance of the lines AC and BC and of the point C is understood, an examination may be made of the various areas in Fig. 171. In the area above the curve ACB liquid only can exist, since the

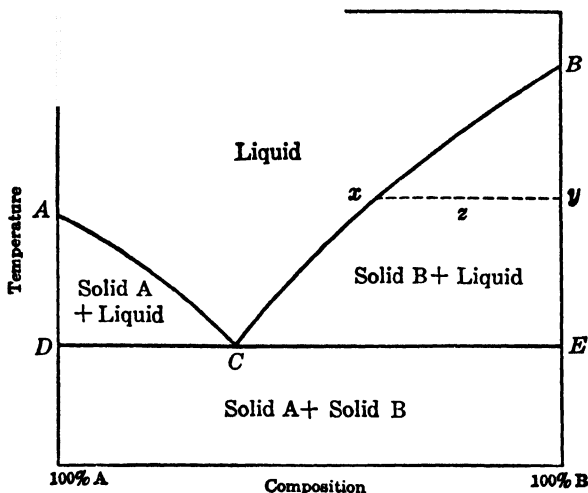


FIG. 171. Solid-liquid equilibria

temperature is above the melting point of any mixture; the condensed systems in this area are bivariant. Since there is one phase only, namely liquid, and two components,

$$F = C - P + 2 = 3 - 1 + 2 = 3,$$

there are three degrees of freedom, of which the pressure is fixed, so that two remain variable. It is necessary, therefore, to state both temperature and composition to define the system exactly; this is in agreement with the representation of such systems by an area in the temperature-composition diagram on which two coordinates are required to define a point. Below the line DCE , passing through the eutectic point C , only solid can exist, since liquid cannot be obtained below the eutectic temperature. In this region there are two solid phases, pure A and pure B. According to the phase rule the condensed system is univariant; the composition of *each phase** is fixed, each being a pure component, and so the only variable is the temperature.

The significance of the areas ADC and BEC must now be considered. At a temperature t , for example, there is equilibrium between the liquid of composition x and the solid represented by y , which is, of course, pure B. Any point between x and y represents a mixture of liquid x and solid y in equilibrium in various proportions at temperature t ; if the point is

* The phase rule is only concerned with the proportions of the components in *each phase*; if the phase is a pure component then its composition is not a variable.

nearer x then there is more liquid and less solid, whereas if it is nearer y there is relatively more solid. At z , for example, the proportion of liquid to solid is in the ratio of the distances yz to xz . Any point in the area BEC represents equilibrium between solid B and liquid, and in the area ADC the equilibrium is between solid A and liquid. The composition of the phases in equilibrium may be obtained by drawing a horizontal line through the particular point, and finding where it cuts AD or BE to give the nature of the solid, and AC or BC to give the composition of the liquid. The line xy , and in fact any horizontal line in these two areas, is a tie-line connecting the compositions of the solid and liquid phases in equilibrium. The curve ACB , called the **liquidus curve**, thus gives the compositions of the liquid phases in equilibrium with the solid phases represented by $ADEB$, which is the **solidus curve**. The two curves meet at the eutectic C , and hence at this temperature the composition of the solid is the same as that of the liquid from which it separates. The equality of composition of the two phases, together with the fact that the eutectic occurs at a definite temperature, led at one time to the erroneous impression that the eutectic system was a chemical compound of the two components (cf. p. 772). For the areas under consideration the phase rule indicates one degree of freedom; this is in agreement with experiment, for if the temperature is fixed the tie-line immediately gives the composition of the liquid and the nature of the solid.

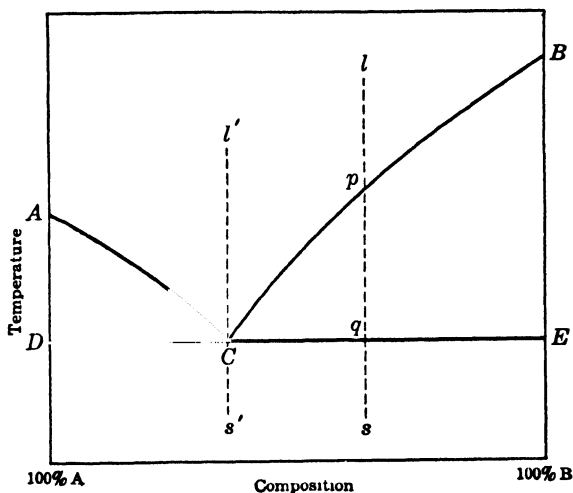


FIG. 172. Cooling of liquid mixtures

Cooling of Liquid Mixtures.—By means of the equilibrium diagram it is possible to forecast the behavior of any system on heating or cooling; this is of importance especially in the study of alloys in metallurgy. Consider, for example, a system of composition represented by the line ls (Fig. 172); l represents the system above its melting point, when it is a

single liquid phase, and s the same system completely solidified. If the liquid l is cooled no solid will separate until the point p on the freezing-point curve BC is reached; at this temperature solid B will commence to deposit. The formation of solid will result in the liberation of the heat of fusion, so that at the point p the rate of cooling will be checked. As the temperature continues to fall, the state of the system as a whole will be indicated by points between p and q , and as shown above these represent solid B and liquid in equilibrium, the composition of the latter changing from p to C , as the temperature falls from p to q . When the eutectic point C is reached, the second solid, i.e., A , commences to deposit, and now the temperature must remain constant until all the liquid has solidified, since the phase rule shows there is only one temperature at which the liquid can be in equilibrium with two solid phases in a condensed system. The attainment of the eutectic temperature will be marked, therefore, by a complete arrest in the rate of cooling, and only when the whole system has solidified can the temperature fall further, e.g., from q to s .

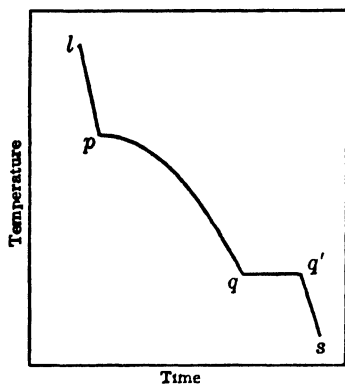


FIG. 173. Cooling curve

If the temperature of the system, starting with the molten liquid, were plotted against time a curve such as shown in Fig. 173 is to be expected; the letters correspond to those in Fig. 172. From l to p the liquid cools rapidly, then at p solid commences to separate and the heat evolved results in a slowing down in the rate of cooling. The composition of the liquid phase changes, as solid B deposits continuously between p and q , and the temperature falls slowly. At q , the eutectic point, the temperature remains constant while both solids A and B separate out, and only when solidification is complete, at q' , does the tem-

perature fall again to s . This type of cooling curve will be obtained for any system to the right or left of C in Fig. 172; the only difference is that if it is to the left solid A will separate out first, whereas to the right B is the first solid to deposit. The temperature p will, of course, depend on the initial composition of the mixture, but that of qq' will be constant throughout. It will be evident that the closer the composition of the system is to that of the eutectic, the shorter will be the portion pq and the longer the halt qq' at the eutectic temperature. If the composition of the liquid mixture coincides exactly with that of the eutectic, as at l' in Fig. 172, then on cooling solid will only separate when the eutectic point is reached at C ; the system will now solidify completely. The cooling curve will show one break only as in Fig. 174, where cc' is at the eutectic temperature. At c' no more liquid remains, and the solid can then cool from c' to s' , or lower.

Thermal Analysis.—The anticipations discussed above are in agreement with observation, and they form the basis of the method of **thermal analysis** for the study of solid-liquid equilibria. A series of mixtures, of known composition, of the two components are made up and heated until a homogeneous liquid is formed; each is allowed to cool at a uniform rate and the temperature is recorded at regular intervals so as to obtain the cooling curves. The first halt in the curve, corresponding to p , occurs at the freezing point, which varies with the composition of the mixture; the second at q gives the eutectic temperature and for a system of the type under consideration should be constant. In this way the equilibrium, or phase, diagrams can be plotted for the whole system as in Figs. 171 and 172; the temperatures of the first halt give the curves AC and BC and the second should give the line DCE . In order to complete the diagram it is necessary to have the freezing, or melting, points of pure A and B; these can also be determined by the methods of thermal analysis, for the pure liquid will solidify at a constant temperature.

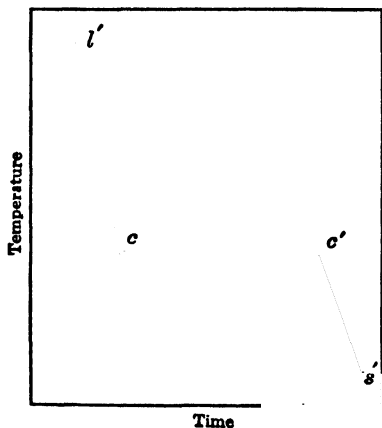


FIG. 174. Cooling curve of eutectic mixture

With some liquid mixtures the separation of solid phase does not occur readily and supercooling is observed; instead of the break in the cooling curve at p , the cooling continues steadily along a continuation of lp and then may rise suddenly to meet the line pq which it subsequently follows (Fig. 175). The correct freezing point may then be obtained by extrapolation back to p , as shown. To avoid supercooling the melt may be "seeded" by addition of crystals of the substance which should separate (cf. p. 647); by the provision of crystallization nuclei the separation of solid is generally encouraged.

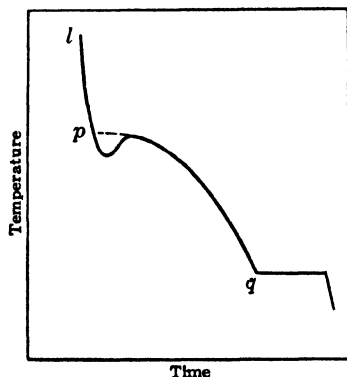


FIG. 175. Cooling curve with supercooling

The Thaw-Melt Method.—In many cases the thermal analysis procedure is not altogether satisfactory, especially when supercooling is serious. Even if there is no supercooling, the temperature arrest at the eutectic point is often difficult to observe on cooling, for the system contains a large proportion of solid and stirring is difficult; solid also adheres to the thermometer and so the latter does not record the equilibrium temperature. For these and other reasons the thaw-melt method of H. Rheinboldt (1925) has attracted interest; *

* A somewhat similar procedure was used by A. Stock (1909).

it involves heating, instead of cooling, the mixture of known composition. If in Fig. 172, for example, the temperature of the solid represented by *s* is gently raised, no change will occur until the point *q*, the eutectic temperature, is reached; liquid of composition *C* will then commence to form. As the temperature is raised from *q* to *p* the quantity of liquid increases and its composition changes from *C* to *p*; at *p* the last trace of solid is about to disappear and above *p* the system is completely liquid. In the thaw-melt method, a well mixed sample—made by melting, cooling and then grinding—containing known amounts of the two solids is heated slowly. A wide capillary tube, similar to those used for determining melting points in the laboratory, may be used, the fine powder being pressed well down with a thin rod. The temperature at which the first minute drops of liquid appear, i.e., at *q*, the so-called “thaw point,” is the eutectic for the given system, and that at which the last trace of solid just disappears is the “melt point,” equivalent to *p* in Fig. 172, giving the appropriate point on the curves *AC* and *BC*. The thaw-melt procedure incidentally emphasizes the fact that an impure substance may melt over a considerable range of temperature. Only a pure substance, or a mixture having the composition of the eutectic, melts sharply at a definite temperature.²⁰

Examples.—Many instances are known of systems having two components completely miscible in the liquid but immiscible in the solid state, so that each separates out in the pure form when the liquid is cooled. The two components may be metals, organic compounds, salts or a salt and water. The last category is a special one which will be considered in more detail later (p. 771); since salts either decompose before melting or melt at high temperatures, the complete equilibrium curve can be obtained in but a few cases, and then only by working under relatively high pressure to prevent complete evaporation of the water. Some examples of two-component systems of the type under discussion are given in Table 129.

TABLE 129. SYSTEMS WITH SIMPLE EUTECTIC POINTS

A	M. Pt.	B	M. Pt.	Eutectic
Antimony	630° c.	Lead	326° c.	246° c.
Silicon	1412°	Aluminum	657°	578°
Bismuth	317°	Cadmium	268°	146°
Potassium chloride	790°	Silver chloride	451°	306°
α -Sodium sulfate	881°	Sodium chloride	797°	623°
<i>o</i> -Nitrophenol	44.1°	<i>p</i> -Toluidine	43.3°	15.6°
Benzene	5.4°	Methyl chloride	-63.5°	-79°
<i>d</i> -Pinene	-64°	<i>l</i> -Pinene	-64°	-120°

Transition Points.—If one of the components has two enantiotropic crystalline forms (p. 469) the equilibrium diagram becomes slightly more complicated, as in Fig. 176, which would be obtained if B existed in two solid forms designated α and β . The curve *AC* is quite normal and requires no further discussion. The curve *BC* is divided into two parts *BF* and *FC* meeting at *F*; this is the transition point, and at temperatures above *FG* the α -form is stable, whereas below it the β - is the stable form. At *F* both solid forms can exist in equilibrium with liquid, and hence

according to the phase rule it represents an invariant point at a given pressure; there is thus only one temperature at which this equilibrium can occur. Along the line BF the solid α -form separates out on cooling, and along FC the solid in equilibrium with liquid is the β -form. Points in the area BFG represent mixtures of solid $B(\alpha)$ and liquid, whereas in the area $GFCE$ the systems represent solid $B(\beta)$ and liquid. If the liquid l is cooled the temperature will fall rapidly until p is reached, then solid $B(\alpha)$ commences to deposit and there is a break in the cooling curve. Separation of $B(\alpha)$ continues until the transition point is reached; at this temperature $B(\beta)$ commences to form, but as long as both α and β are present the temperature must remain constant. The fall of temperature

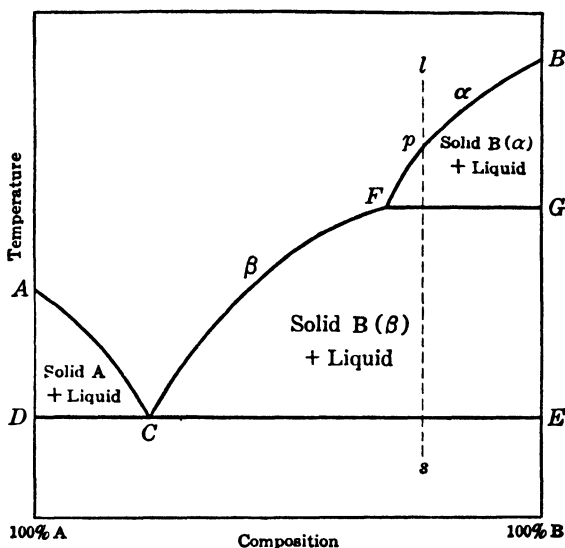
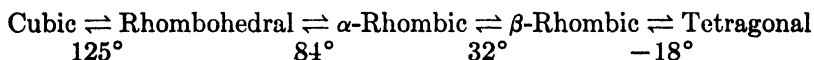


FIG. 176. System with transition point

is thus completely arrested while the α -form changes over into the β -form; when this is complete cooling once more sets in, with $B(\beta)$ being deposited, along FC until the eutectic point C is reached when solids A and $B(\beta)$ form an invariant system with liquid. The temperature again remains constant until the whole of the liquid has solidified. The cooling curve will thus show (i) a break dependent on the composition of the system, (ii) a horizontal portion at the transition point, and (iii) a second horizontal portion at the eutectic temperature. The closer the point p is to F the shorter will be the portion (ii), since there is relatively less $B(\alpha)$ formed; if p falls to the left of F , then $B(\beta)$ separates out on cooling, and the system behaves quite normally. Solid carbon tetrachloride exists in two forms, with a transition point at -48°C ., and all systems involving this substance which have been examined show a break in the equilibrium

curve at this temperature, provided, of course, that it is not below the eutectic point of the system. If the latter were the case the curves would be of the simple type, for the transition would only occur after the whole had solidified. Ammonium nitrate melts at 169.9°C . and has the following transition points:



The curve for the equilibrium between solid ammonium nitrate and liquid mixtures with water, i.e., aqueous solutions, shows breaks at these temperatures.

Pseudo-Binary Systems.—An interesting type of two-component system occurs when the two substances are tautomeric, that is they exhibit dynamic isomerism; each form is capable of conversion into the other and at each temperature there is an equilibrium system at which the two forms, α and β , are present in definite proportions. If the $\alpha \rightleftharpoons \beta$ equilibrium is attained slowly, then it is possible to obtain a solid-liquid phase diagram of the simple type described above, each isomeride acting as a single individual (Fig. 177); the

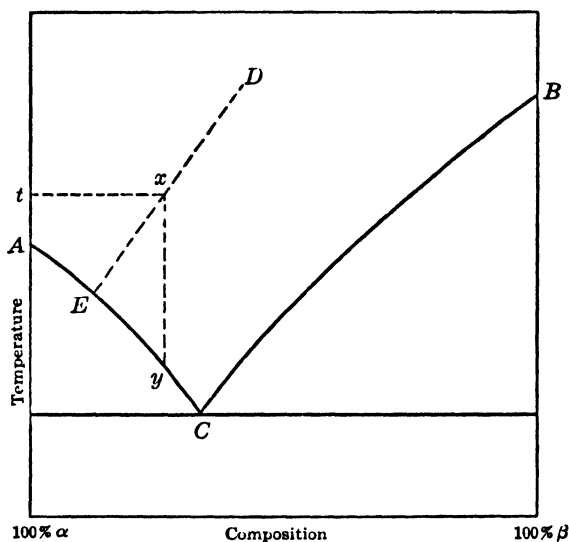


FIG. 177. Pseudo-binary system

curves AC and BC , however, represent metastable equilibrium conditions. When true equilibrium between the isomerides is attained the system will always have the same ratio of α to β at a given temperature, irrespective of the amounts at the commencement. The broken line DE , for example, might represent the true $\alpha \rightleftharpoons \beta$ equilibrium in the liquid phase; the equilibrium system is seen to contain, in this hypothetical instance, relatively more of the α -isomeride, the amount increasing as the temperature falls. Such a curve would

mean, according to the Le Chatelier principle (p. 831), that the conversion of α to β is accompanied by the absorption of heat, and *vice versa*. If instead of cooling the liquid rapidly, which would be necessary to obtain the curves AC and BC , it were cooled very slowly, so as to permit the establishment of the tautomeric equilibrium, the composition would always be on the line DE . As the system cooled it would follow this line, and at the point E solid would separate; if sufficient time were allowed the whole system would solidify at this point, the solid having the same equilibrium composition as the liquid at that temperature. The point E is called the "natural," or "equilibrium," freezing point of the tautomeric substance. In time every liquid-solid system must reach the point E , for this represents the only true state of equilibrium between liquid and solid. If tautomeric equilibrium is attained rapidly then it may be the only point actually observable, for then the curves AC and BC could not be obtained. For systems in which the $\alpha \rightleftharpoons \beta$ transformation occurs slowly the freezing-point diagram may be used to study the proportions of the two dynamic isomerides at equilibrium at various temperatures. For example, if at temperature t the system is allowed to come to equilibrium, its composition will be given by the point x ; if now it is cooled rapidly solid will commence to separate at y . If the temperature is noted the composition of x can be read off from the freezing-point diagram which must, of course, be prepared from measurements on a number of mixtures of known composition. Systems consisting of two mutually convertible polymers, e.g., acetaldehyde and paraldehyde, behave in an analogous manner. Cases of pseudo-binary systems involving dynamic isomerides have been studied by W. D. Bancroft (1898) and by others; examples are the α - and β -benzaloximes, the α - and β -acetaldoximes, and ammonium thiocyanate and thiourea. It is of interest that in these cases the line DE is vertical, indicating that the mutual transformation of the tautomerides is not accompanied by any appreciable heat change.²¹

I. B. The Components are only Partially Miscible in the Liquid State and the Solid Phases Consist of Pure Components.—It has been assumed hitherto that the components are completely miscible in the liquid state, but this is not always the case. Over a certain range of concentration the liquids may be partially miscible and two liquid phases are formed. The type of equilibrium diagram obtained is shown in Fig. 178. Starting with pure A and adding increasing amounts of B, the freezing point falls along AF , solid A separating; at F the solubility limit of B in liquid A is reached and two liquid layers, represented by the points F and G , are formed. There are now three phases, two liquid and one solid, in equilibrium, and hence the phase rule requires the condensed system to be invariant, since the pressure is fixed. Addition of more B means, therefore, that the temperature will remain constant as long as solid A is in equilibrium with the two liquids F and G . The relative amounts of the two layers alter, as the composition of the system as a whole changes from F to G , the proportion of the layer G increasing and of F decreasing as the concentration of B is increased. Eventually, when the composition of the whole is represented by G , there is again only one liquid layer, and the system becomes univariant once more, solid A separating along GC until the eutectic C is reached when the whole solidifies. The curve FMG is

the ordinary solubility curve for two partially miscible liquids (p. 724); there are two liquid phases only, and the condensed system is univariant, so that at each temperature the compositions of the phases in equilibrium are fixed by a horizontal tie-line. If the temperature of the system FG is raised, solid A will disappear and the compositions of the layers will follow the curve FMG , where M is the critical solution temperature.

The properties of the system can best be illustrated by considering the changes observed on cooling liquid mixtures. If the homogeneous liquid l is cooled, it commences to form two layers at p , the compositions of which are p and p' ; with further cooling the compositions change along pF and $p'G$ respectively. When the system as a whole is represented by

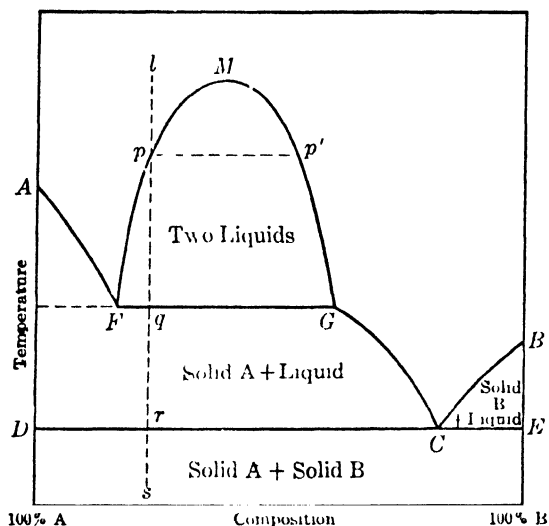


FIG. 178 Partial miscibility of liquid phases

q , the liquid layers are F and G ; then solid A commences to separate and the temperature remains constant as the system is univariant. As A continues to deposit, the relative amount of the liquid layer richer in B, i.e., layer G , increases, and eventually layer F disappears; then the temperature can fall again. From q to r the behavior is like that for a simple two-component system, with solid A in equilibrium with one liquid layer, the composition of which changes along the line GC . At the eutectic point the other solid (B) deposits, and the temperature again remains constant until no liquid is left. The mixture of solid A and B can then cool along rs . Any liquid whose composition is such that the point equivalent to p falls on the curve FMG behaves in exactly the same manner. If the initial system is given by a point above AF , then solid A will separate from the one-phase liquid, and the temperature will fall steadily until the

composition of the liquid reaches F , when the second layer G will be formed. Further cooling results, as before, in the deposition of more solid A until the layer F has disappeared and only G remains; subsequent cooling leads to further deposition of A , the liquid composition changing along GC , as already described. Liquid systems lying above GC and BC deposit solid A or solid B , respectively, on cooling and behave normally without any separation into two liquid layers. Examples of systems forming two liquid layers are given in Table 130; the invariant temperatures, at which

TABLE 130. SOLID-LIQUID SYSTEMS WITH TWO LIQUID LAYERS

Component A	M. Pt.	Component B	M. Pt.	Invariant Point
Benzoic acid	121° c.	Water	0° c.	98° c
Phenol	43°	Water	0°	1.7°
Succinic nitrile	54°	Water	0°	18.5°
Resorcinol	110°	Benzene	5°	95.5°
Zinc	418°	Bismuth	269°	415°

the two liquids are in equilibrium with solid, are given in the last column. The results given above account for the familiar phenomenon of "melting under water," in which a solid is observed to liquefy at a temperature well below its true melting point. For example, if phenol is added to water, two liquid layers can form at temperatures down to 1.7°; the apparent melting is really the formation of a saturated liquid solution of water in phenol.

A few cases have been reported of the equilibrium between a solid phase and two liquids whose mutual solubilities increase with decreasing temperature; these are somewhat unusual, as will be evident from the nature of the components, viz., potassium iodide and sulfur dioxide, and pyridine and potassium thiocyanate.

II. A. The Two Components Form a Compound with a Congruent Melting Point.—When compound formation occurs between two components, then if the compound is stable in the liquid form, it is improbable that the system will show sufficient positive deviation from ideal behavior for two liquid phases to coexist; hence one liquid only need be envisaged. A few cases of two liquid layers have been reported where solid compound formation occurs between metals, but the compounds are probably of the "lattice" type (p. 397) and are not present in the liquid. If the components A and B form a solid compound AB , then the solid-liquid equilibrium diagram, obtained by either of the methods already described, will be of the type shown in Fig. 179. In addition to the two branches AC and BE , which represent the compositions of liquid in equilibrium with solid A and solid B , respectively, at different temperatures, there is a central portion CDE rising to a maximum. This portion of the curve represents liquid systems in equilibrium with the solid compound AB . It has been found experimentally, in accordance with theoretical requirements, that the maximum D of the curve occurs at the composition of the solid compound; as in this instance it is supposed to contain equimolar

amounts of A and B, the point D' is midway between pure A and pure B. At the point D , therefore, liquid and solid phases have the same composition, so that D may be regarded as the melting point of the compound. In this case the compound AB is said to have a **congruent melting point**, since solid and liquid of the same composition can coexist. It will be noted that at this point the two-component system has virtually become a one-component system, i.e., AB, so that the temperature at D is just as definite as the melting points of the pure components. The melting point of the compound may be above, below or between those of the two components.

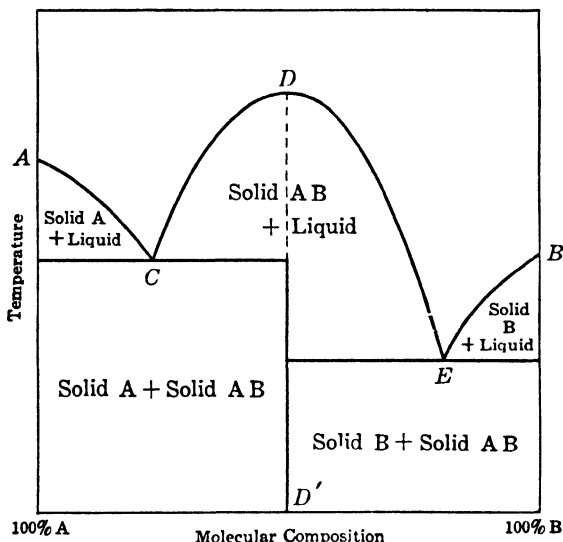


FIG. 179. Compound with congruent melting point

In Fig. 179 it will be seen that there are now two eutectic points, viz., C and E ; at the former, the solids A and AB separate, whereas at the latter the solids are B and AB. The details of the phase diagram can best be understood by regarding it as two diagrams of the simple eutectic type (Fig. 171) placed side by side. To the left of the dividing line DD' the figure gives the conditions of phase equilibrium of the two-component system A and AB, whereas to the right of the line the components may be regarded as AB and B. The details in each part of the diagram correspond exactly with those in Fig. 171, and bearing these facts in mind it is quite simple to determine the behavior of any liquid on cooling, or of any mixture of solids when heated.

When the two components form more than one compound the equilibrium diagram will have a curve analogous to CDE for each compound; for example, Fig. 180 shows the type of curve to be expected for a system in which the compounds A_2B and AB_2 are formed with congruent

melting points; their compositions and melting points are given by D and F , respectively. As in the previous example, the diagram can be readily interpreted by considering it as made up of three separate diagrams for the systems A and A_2B , A_2B and AB_2 , and AB_2 and B , respectively.

If, as suggested, the complete phase diagram is equivalent to separate two-component diagrams placed side by side, then it would be expected that the curves CD and DE , for example in Fig. 180, would intersect sharply instead of merging into each other in the manner shown. It is generally considered that if the compound were perfectly stable at its melting point, that is it melted without decomposition, then the curves

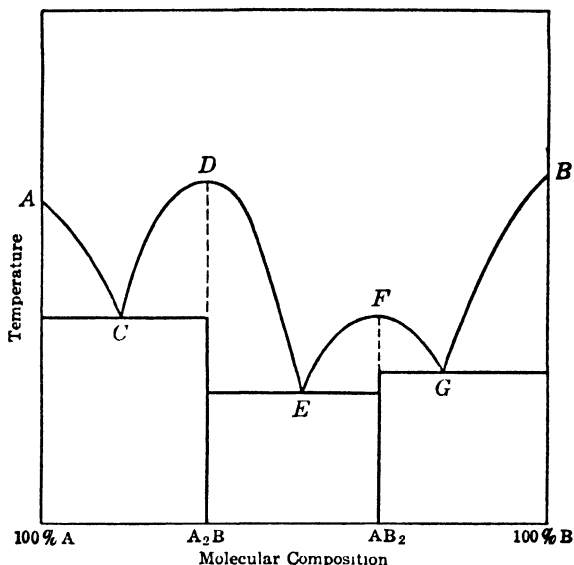
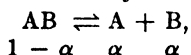


FIG. 180. Formation of two compounds

would show a sharp, instead of a rounded, maximum. Actually some decomposition of the compound into its constituents nearly always occurs; the presence of products of dissociation in the liquid phase depresses the hypothetical melting point the compound would have if there had been no decomposition, and so the curve is flattened. It was suggested by W. D. Bancroft (1898) that the extent of dissociation of the compound into its constituents could be estimated from the flattening, and the method has been worked out by R. Kremann (1904) and J. J. van Laar (1906).

Suppose α is the degree of dissociation of the compound AB at its melting point, then instead of 1 mole of AB the system will be represented by



so that the mole fractions of AB, A and B in the liquid are $(1 - \alpha)/(1 + \alpha)$, $\alpha/(1 + \alpha)$ and $\alpha/(1 + \alpha)$, respectively. Assuming the liquid system to be ideal, the condition of equilibrium (p. 822) requires that

$$K = \frac{x_A \times x_B}{x_{AB}} = \left(\frac{\alpha}{1 + \alpha} \right)^2 \bigg/ \left(\frac{1 - \alpha}{1 + \alpha} \right). \quad (81)$$

If n moles of A are now added to 1 mole of AB, the extent of dissociation when melted will be reduced to α_1 , but as the total number of moles is now $1 + \alpha_1 + n$, the law of equilibrium gives

$$K = \left(\frac{\alpha_1 + n}{1 + \alpha_1 + n} \right) \left(\frac{\alpha_1}{1 + \alpha_1 + n} \right) \bigg/ \left(\frac{1 - \alpha_1}{1 + \alpha_1 + n} \right). \quad (82)$$

Assuming that Raoult's law applies to solutions of A and B in AB, (80) can be written in the form

$$\ln x_{AB} = - \frac{L_f}{R} \left(\frac{1}{T} - \frac{1}{T_0} \right), \quad (80a)$$

where x_{AB} is the mole fraction of AB, L_f is its molar heat of fusion, T_0 its *theoretical* freezing point, and T the actual freezing point of the given solution. The value of L_f may be determined by measuring the depression of the freezing point of AB when an inert substance, which dissolves in it but does not react, is added. In order to obtain α from the actual freezing-point curve a method of trial and error has to be adopted. First, a value for α is assumed and by means of (81) the corresponding equilibrium constant K is calculated; the value for x_{AB} , which is $(1 - \alpha)/(1 + \alpha)$, is then inserted in (80a), and the appropriate T_0 determined. Assuming the same K the degree of dissociation α_1 in the presence of various amounts n of added A can be evaluated; from this the corresponding mole fractions of x_{AB} can be obtained, and by means of (80a) the freezing points T may be calculated. If the value of α has been correctly chosen these points would correspond to the points on the phase diagram, for the portion *DC* of Fig. 179, for example, gives the effect of the addition of A on the freezing point of AB. It is improbable that the correct value of α would have been guessed, and so the procedure is repeated until the calculated freezing-point curve agrees with the experimental one. In this manner the degree of dissociation α of the pure compound at its melting point is obtained. An approximate and much simpler method of computation was proposed by J. J. van Laar (1906), but it appears to give values of α which are too high. The equation deduced was

$$\alpha = \frac{RT^2}{L_f} \cdot \frac{x^2(1 + x)}{4(\Delta T)}, \quad (83)$$

where ΔT is the freezing-point depression of AB resulting from the addition of x mole fraction of either A or B. The value of L_f may be obtained by the approximate relationship $dT/dx = RT^2/L_f$ [equation (35), p. 644], T being the observed freezing point of AB.

It will be noted that in the calculation described above it is assumed that the same value of K applies in (81) and (82). Since the former refers to AB alone and the latter to a mixture of AB and excess of A, the temperatures are different, and so to assume that K is the same for both implies that the heat of dissociation, or of formation, of the compound is zero. The same assumption is involved in other parts of the argument, but it will be sufficient to indicate this

one limitation, which also applies to the derivation of (83). Incidentally, it will be noted, it has been postulated that the systems behave ideally, although this is unlikely to be so in practice.²²

Many examples are to be found of binary systems in which solid compounds occur. The plotting of the freezing-point diagram is the recognized method used to determine whether two components combine to form a stable solid compound or not, and if they do, to indicate its composition.* Instances of systems in which one or more solid compounds are formed are given in Table 131.

TABLE 131. BINARY SYSTEMS WITH STABLE COMPOUNDS

A	M. Pt.	B	M. Pt.	Compound	M. Pt.
Aluminum	657° c.	Magnesium	650° c.	A ₃ B ₄	463° c.
Gold	1064°	Tin	232°	AB	425°
Calcium chloride	777°	Potassium chloride	776°	AB	754°
Sodium fluoride	986°	α -Sodium sulfate	881°	AB	781°
Diphenylamine	52.8°	Benzophenone	47.7°	AB	40.2°
Urea	132°	Phenol	43°	AB ₂	61°

II. B. The Two Components Form a Compound with an Incongruent Melting Point.-- Sometimes the compound formed is so unstable that it decomposes completely at a temperature below its melting point, so that the solid cannot be in equilibrium with a liquid having the same composition as itself; in other words, it has no true melting point. The phase-equilibrium diagram for this type of behavior is shown in Fig. 181, in which it is supposed that a 1 : 2 compound AB₂ is formed. At a temperature *E*, below the hypothetical melting point *D*, the compound AB₂ dissociates completely into its constituents; along *CE* the solid AB₂ separates, but along *EB* the component B forms the solid phase. The point *E* is the **incongruent melting point** of the compound. If a liquid, such as that represented by *l*, is cooled the first solid to separate will be B; then when the temperature reaches *E* the compound AB₂ commences to form. Since there are now one liquid and two solid phases the condensed system is invariant, and the temperature at *E* will remain constant until the solid B has been completely converted into solid compound, or the liquid phase disappears. For all points to the right of *D* the latter alternative will occur, because the original system is richer in B than is the compound, and there is insufficient A in the liquid to convert all the solid B into AB₂. When the composition of the system lies between *D* and *E*, the solid which has separated will be completely changed into AB₂, and then the temperature will fall again until the eutectic point *C* is reached.

It will be evident that, as far as the curves are concerned, Fig. 181 bears a definite resemblance to Fig. 176, where there is no compound

* The absence of a solid compound does not dispose of the possibility that combination may occur in the liquid phase. Alternatively, the existence of a solid compound does not prove that the same combination exists in the liquid; the former may be the result of a particular type of packing in the crystal, as in many intermetallic compounds (see Chap. V).

formation but the component B exists in two modifications. The point *E* in Fig. 181, like *F* in Fig. 176, is also frequently called a transition point, especially when the components are a salt and water and the compound is a salt hydrate (see p. 773). It would be possible to distinguish between the two types of behavior by examination of the solid phases, but a simpler method is to examine the cooling curves. When no compound is formed, all mixtures which deposit solid between *B* and *F* (Fig. 176) show two complete arrests on cooling: one at the transition point and the other at the eutectic temperature. If an incongruently melting compound is formed, however, there is only one complete arrest if the system contains more B than the compound, and this occurs at the transition point *E*; for mixtures between *D* and *E* (Fig. 181) there are two arrests, at *E* and *C*,

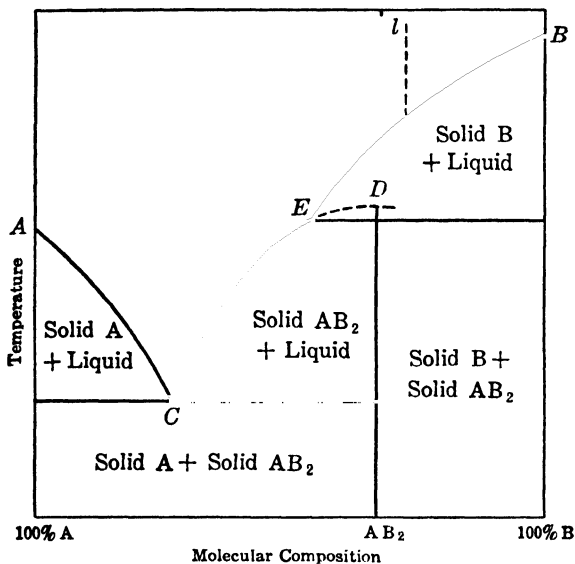


FIG. 181. Compound with incongruent melting point

respectively. The formula of the compound can also be obtained without analysis, for it will correspond in composition with the system giving the longest halt at the transition point, for a given weight of material. Another possibility is to study by the thaw-melt procedure the behavior of solid mixtures made by melting the components together and then cooling the liquid. Examination of the various areas in Fig. 181 will show the difference to be expected for mixtures containing more A or more B, respectively, than the compound.

Systems in which compounds having incongruent melting points are formed are gold-antimony ($AuSb_2$), magnesium-nickel (Mg_2Ni ; $MgNi_2$ has a congruent m. pt.), sodium-bismuth ($NaBi$; Na_3Bi has a congruent m. pt.), potassium chloride-cupric chloride ($2KCl \cdot CuCl_2$), potassium

sulfate-cadmium sulfate (AB_3 and AB_2), picric acid-benzene (AB), and acetamide-salicylic acid (AB).

III. The Two Components Form a Continuous Series of Solid Solutions.—When the solid phase is not a pure substance but a solid solution, it is possible (p. 651) for the freezing point of one component to be *raised* by the addition of another. Further, if the two solids are completely soluble in each other, only one solid phase can exist, for a homogeneous solid solution constitutes a single solid phase, and only one liquid phase can form; the existence of solid solutions, as might be anticipated, is not accompanied by partial miscibility in the liquid phase. Since the maximum number of phases is two, for the condensed system, the minimum number of degrees of freedom, including pressure, is two; hence an invariant system is impossible and there will be no singular point, i.e., discontinuity, on the phase diagram. The freezing-point diagrams obtained fall into three groups which are analogous to the three main categories of boiling-point curves (p. 716); the freezing points of all systems may lie between those of the pure components, or there may be a maximum, or a minimum.

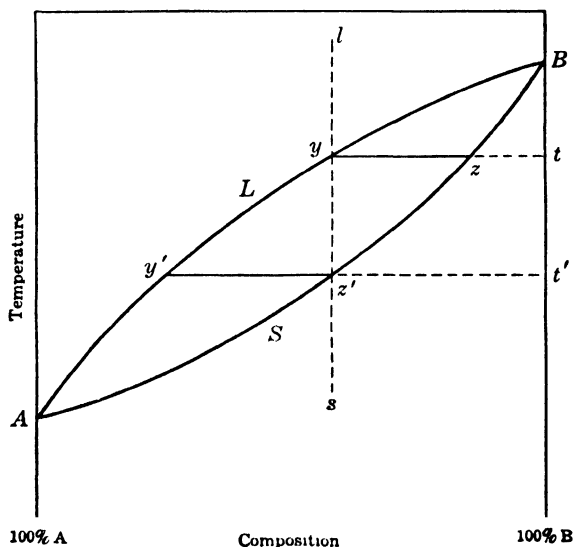


FIG. 182. Continuous series of solid solutions

(i) *The Freezing-Point Curve has Neither Maximum nor Minimum.*—A typical equilibrium diagram is shown in Fig. 182; the upper, or **liquidus**, curve (L) represents the composition of the liquid phase in equilibrium with solid whose composition is given by the lower, or **solidus**, curve (S). It is seen that the composition of the solid phase changes continuously with that of the liquid from which it separates. A rule, similar to that of

Konowaloff for liquid-vapor systems, applies here: at any temperature the liquid phase contains relatively more than the solid phase of that component by the addition of which the freezing point is depressed. If the liquid l is cooled, freezing will commence at y and the composition of the solid phase separating will be given by z ; in this case the addition of A lowers the freezing point, and so the liquid contains relatively more of this component than does the solid. As the latter contains relatively more B , the liquid will become richer in A , and so the temperature follows the curve ya , solid separating the whole time. The state of the system as a whole follows the line ls , the compositions of the liquid and solid solutions in equilibrium at any instant being given by a horizontal tie-line through the point. As the temperature falls the amount of solid increases relative to that of the liquid, and when the system reaches the point z' the last traces of liquid y' are just on the point of disappearing; at this temperature solidification is just complete. Strictly speaking, this result is only obtained if the solid and liquid phases remain in equilibrium all the time as cooling proceeds; in the treatment here this will be assumed, although it is not always the case.* Separation of solid thus commences at y and is complete at z' ; for this reason the liquidus curve is sometimes called the freezing-point curve, and the solidus curve is called the melting-point curve. All systems above the liquidus curve represent a single liquid phase, whereas all below the solidus curve are a single solid phase; in

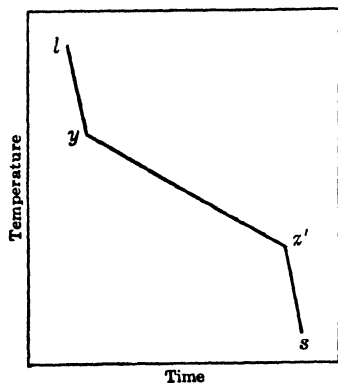


FIG. 183. Cooling curve with formation of solid solution

between the two curves points represent liquid and solid solutions in equilibrium.

From what has been stated it will be apparent that on thermal analysis mixtures of the type being considered will give a cooling curve such as that in Fig. 183; solid commences to separate at y and solidification is complete at z' , thus giving the appropriate points on liquidus and solidus curves, respectively. There is no invariant point and so there is no complete arrest, although the rate of cooling is less while solid is separating than when liquid only (ly) or when solid only ($z's$) is present. If a solid solution is heated, e.g., from s upward, as in the thaw-melt method, the appearance of

the first drops of liquid gives the point on the solidus curve (z') for the mixture of the given composition, and the temperature at which liquefaction is complete indicates the position on the liquidus curve (y). The data necessary to plot the complete equilibrium diagram may thus be obtained by either procedure.

* In practice the final solidification temperature is generally lower than if equilibrium were established.

When two components do not form solid solutions, separation of both components by fractional crystallization is not possible; one or other of them may crystallize out first and be removed, but both cannot be obtained from one mixture. When the system gives a continuous series of solid solutions, however, and the freezing points of all mixtures lie between those of the pure components, separation by fractional crystallization is possible, just as fractional distillation is possible in the analogous liquid-vapor system. If the liquid l (Fig. 184) is cooled down to x , it will consist of a mixture of almost equal parts of liquid a and solid b ; the former is appreciably richer in A and the latter in B than the original mixture. If the layer a is now separated and cooled to x' , there will be about equal amounts of liquid a' and solid b' ; in this way almost pure A will soon be obtained. Similarly, if the solid b is separated and melted and then cooled again to the point x'' , the system will be made up of a mixture of liquid a'' and solid b'' . By repeating this procedure with the solid it is possible to obtain almost pure B after a few recrystallizations.

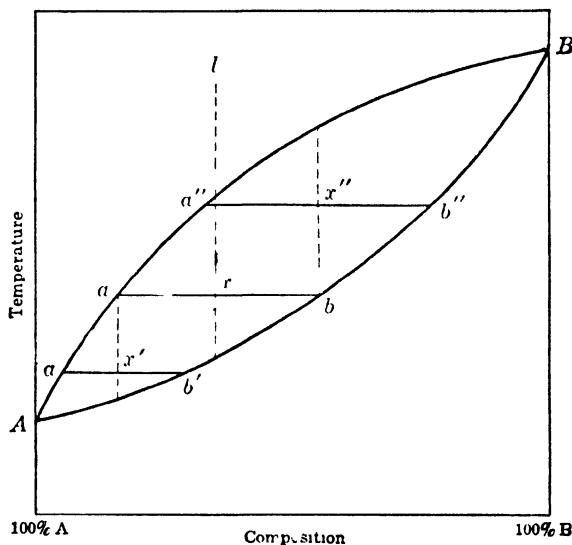


FIG. 184. Fractional crystallization of solid solutions

Substances of similar constitution, or closely related elements, usually form solid solutions; examples are cobalt-nickel, gold-silver, gold-platinum, silver chloride-sodium chloride, lead chloride-lead bromide, naphthalene- β -naphthol, α -monochloro- α -monobromocinnamic aldehyde.

Mixtures of optical isomerides sometimes show unusual behavior; for example, in the systems *d*- and *l*-camphoroxime, *d*- and *l*-borneol, and probably in a few others also, the freezing points of all mixtures are the same as that of the pure components. The solidus and liquidus curves are then horizontal straight lines coincident with each other. Every

liquid mixture of *d*- and *l*-borneol, for example, freezes and every solid solution melts sharply at 207° c. When the two components are *d*-borneol (m. pt. 207° c.) and *d*-camphor (m. pt. 178° c.), a continuous series of solid solutions is formed, and these have the same composition as the liquid from which they separate. The solidus and liquidus curves are thus identical straight lines joining the melting points of the two pure components.

(ii) *The Freezing-Point Curve has a Maximum.*—The type of equilibrium diagram shown in Fig. 185 is not common; it is seen that each

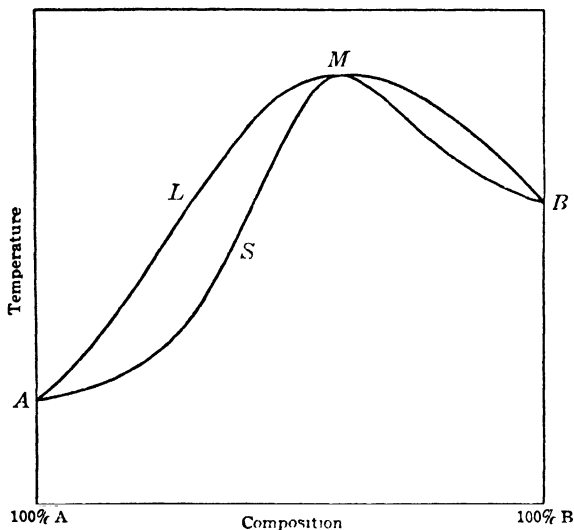


FIG. 185. Solid solutions with maximum melting point

component *raises* the freezing point of the other, and that the liquidus and solidus curves coincide at the maximum *M*. The relative position of the two curves is determined by the rule given above; the liquid always contains a larger proportion of the component which lowers the freezing point. At *M* the liquid and solid in equilibrium have the same composition; the liquid solution will then freeze, and the solid solution melt, sharply at a definite temperature just like a pure substance. Any attempt at separation of the components of the mixture by fractional crystallization will clearly lead to a liquid phase approaching pure A, or pure B, according as the initial composition is between A and M, or between B and M, respectively; the composition of the solid will always tend towards that of *M*.

Very few instances are known of systems yielding a continuous solid solutions and having a freezing-point curve with a maximum. It may be significant that these all involve *d*- and *l*-forms of the same optically active compound, viz., *d*- and *l*-carvoxime, *d*- and *l*-monobornyl esters of malonic acid and, possibly, *d*- and *l*- α -bromocamphor. Since, in each

case, the two components have the same melting point, the freezing point curves are symmetrical, the maximum occurring when equimolar amounts of the two substances are present. It has been suggested that the maximum represents a racemic compound, so that the left-hand side of the equilibrium diagram (Fig. 185) is for the system *d*-form + *dl*-compound, whereas the right-hand side is for the system *dl*-compound + *l*-form. There seems to be no obvious method, however, for proving or refuting this suggestion.

(iii) *The Freezing-Point Curve has a Minimum.*—A number of systems of which the solid components are completely miscible have freezing-point curves with a minimum, as in Fig. 186; here, as in other cases, the liquidus

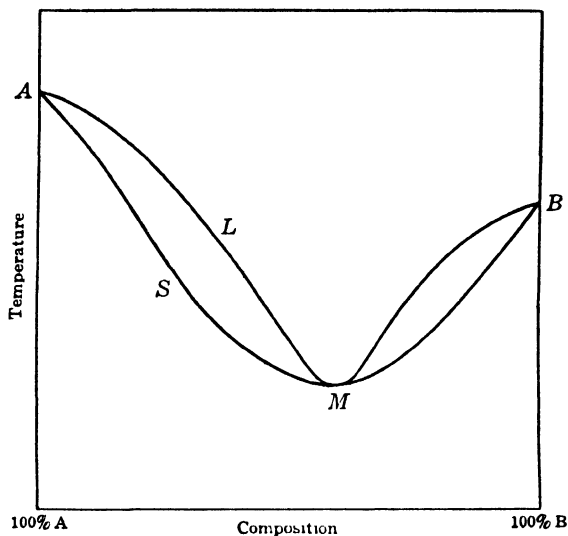


FIG. 186. Solid solutions with minimum melting point

curve lies above the solidus curve, and the two touch at the minimum *M*. At this point solid and liquid in equilibrium have the same composition and so the system will melt and freeze sharply, like a pure substance; the solid is actually a solution of definite composition, but the amounts of A and B are not necessarily in stoichiometric proportions, as would be the case if it were a compound. Without going into details, it will be evident that if a mixture of A and B is fractionally crystallized the solid phase will tend towards pure A or pure B, depending on the composition of the original system, whereas the composition of the liquid will approach that of the mixture of minimum freezing point.

Among systems giving liquid-solid equilibrium diagrams of the type shown in Fig. 186 may be mentioned copper-manganese, copper-gold, manganese-nickel, cobalt-manganese, arsenic-antimony, mercuric bromide-mercuric iodide, potassium chloride-potassium bromide, sodium carbonate-potassium carbonate, and *p*-chloriodo-*p*-dichlorobenzene.

IV. The Two Components are only Partially Miscible in the Solid State.—It frequently happens that two substances can form solid solutions with one another to a limited extent only, so that between definite concentration limits, depending on temperature, two conjugate solid solutions can exist; the phenomenon is quite analogous to that occurring with partially miscible liquids. Solid A can dissolve a certain quantity of solid B to form a single solid solution, and similarly solid B can dissolve a certain amount of A, but if these concentrations are exceeded two solid phases, each of which is a homogeneous solution of A and B, is produced. Since it is now possible to have three phases, two solid and one liquid, the condensed system can be invariant, and so the solid-liquid equilibrium curves will have a discontinuity. Two such points may occur, either a eutectic or a transition point, and as both are found in practice they will be considered in turn.

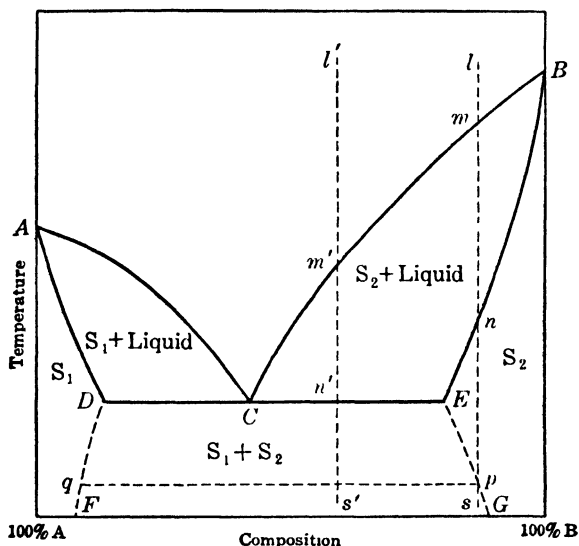


FIG. 187. Partially miscible solid phases with eutectic point

(i) *The System has a Eutectic Point.*—In Fig. 187 AC and BC are the liquidus curves, and AD and BE the corresponding solidus curves; along AD the solids may be regarded as solutions of B in excess of A (S_1 type), and along BE the solid solutions are of A in excess of B (S_2 type). The liquidus curves meet at C which is a eutectic point for the system. It differs in one respect from the ordinary eutectic of Fig. 171; the solid phases in equilibrium with liquid are not pure components, but the two solid solutions represented by the points D and E . This eutectic is, however, a true invariant point for the condensed system, and is the lowest temperature at which liquid can exist. It can be seen that systems giving

curves of the kind shown in Fig. 187 are really in between those for components which are completely immiscible in the solid state (Fig. 171) and those forming a continuous series of solid solutions with a minimum melting point (Fig. 186). If the solids in Fig. 187 are imagined to become more and more miscible, the points D and E will eventually coincide, and the curves will correspond to those of Fig. 186; on the other hand, if the solids become less miscible, D and E will recede from each other and eventually the compositions coincide with pure A and B respectively as in Fig. 171.

In systems of the type under consideration it is of special interest to examine further the mutual equilibrium of the two solid solutions, for the solubilities will vary with temperature, just as with two partially miscible liquids. As a general rule the mutual solubilities of two solids decrease as the temperature is lowered, and the conditions of the solid-solid equilibrium may be indicated by the two curves DF and EG ; theoretically, a critical solution temperature should be found above DE , but in this region liquid only is obtainable and the solids have no real existence. The compositions of the two solid solutions in equilibrium at any temperature may be obtained by drawing a horizontal tie-line from DF to EG at that temperature; for example, the solids p and q will be in equilibrium with one another.

The results to be expected on cooling liquids of different compositions are of particular importance for the study of alloys. If the liquid l is cooled, solid commences to separate at m and, provided the solid solution changes its composition so as to maintain equilibrium with the liquid, solidification is complete at n . So the general behavior, and the form of the cooling curve, will be identical with that for a system giving a continuous series of solid solutions (Fig. 182). When the temperature reaches that of p , however, the solid solution will reach its saturation concentration, and two solid solutions, p and q , will form. Transitions of this type do not occur rapidly in the solid state, and the temperature would have to be maintained at p for some time for the true equilibrium to be established; if the cooling is rapid, a single metastable solid solution may still exist at s , and in fact the lower the temperature the slower the rate of attainment of equilibrium. If time is allowed, however, the compositions of the two solid solutions in equilibrium would change along pG and qF , respectively. The behavior described will apply to any system falling to the right of E or to the left of D .

When the composition of the liquid phase brings the system within the range DE , somewhat different phenomena are observed. For example, if l' is cooled solid solution commences to deposit at m' ; this continues until n' is reached when the liquid is in equilibrium with the two solid solutions D and E . The condensed system is now invariant and the temperature remains constant at the eutectic point until it has completely solidified. On further cooling, the compositions of the two solid solutions will change along DF and EG , provided equilibrium is attained. The cooling curves obtained in thermal analysis will be of the same type as those given by a simple eutectic system (p. 748). If a liquid having the same composition as the eutectic is cooled, it will solidify at a definite temperature, and the cooling curve will show one complete arrest, as

ature remains constant. As the liquid phase *C* contains relatively more of the component *A* than does either of the solid solutions, the liquid does not solidify completely at the transition point, as would be the case for mixtures falling between *D* and *E*, but the solid solution *E* gradually disappears. When it has completely gone the system is univariant and the temperature can fall once more from *n* to *p*, the composition of the liquid changing along *CA* and the solid along *DA*; at the point *p* the whole system has solidified as one solid solution.* The temperature of the solid now falls to *q* when two solid solutions will form if the system is allowed to reach equilibrium.

Systems forming two series of solid solutions with a transition point are cadmium-mercury, silver chloride-lithium chloride, silver nitrate-sodium nitrate, magnesium silicate-manganese silicate, and *p*-iodochloro-*p*-di-iodobenzene.

V. The Two Components are Miscible in the Solid State and also Form a Compound.—For convenience and simplicity of treatment the cases hitherto have each represented one particular type of behavior. There are, however, many complicated examples, particularly in connection with the study of metallic alloys, in which several different possibilities arise in the one binary system. The first case to be considered arises

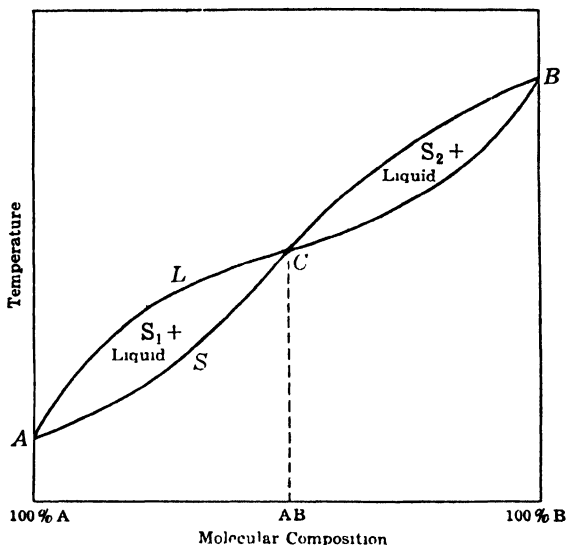


FIG. 189. Solid solutions with compound formation

when the two components form a homogeneous series of solid solutions, and they also form a compound miscible with either component. The

* It is seen that at the temperature of *C* the solid solutions of the S_2 type have been converted into those of the S_1 type; hence the justification of the use of the term transition point. The term peritectic point is used, however, to describe a transition of this type where one solid solution changes into another.

phase diagram in Fig. 189 is then obtained, the point *C*, where liquidus (upper) and solidus (lower) curves touch, giving the formula and melting point of the compound. The curves as a whole are equivalent to two diagrams, one for *A* and *AB*, and the other for *AB* and *B*, placed side by side; in each case there is a continuous series of solid solutions. Examples of this type of behavior are iodine and bromine, and magnesium and cadmium.

In the case cited above the melting point of the compound lies between those of the pure components. If, however, it should be higher than the values for both components, the equilibrium diagram would be similar to that depicted in Fig. 185, provided the compound were completely miscible with each of the components in the solid state. As mentioned on p. 765, the suggestion has been made that the *dl*-systems which give diagrams of the type of Fig. 185 do actually involve compound formation; however, it is not certain that this is the case.

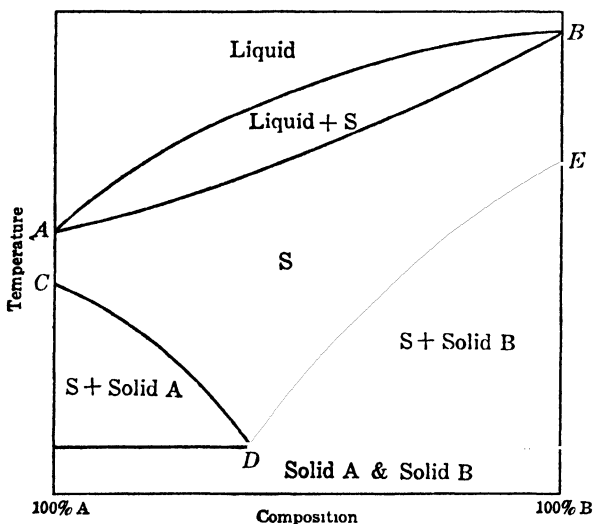


FIG. 190. Changes in the solid state: formation of eutectoid

Another possibility arises when two components which are not completely miscible in the solid state form a compound; an equilibrium diagram equivalent to two sets of curves of the type of Fig. 187 is then obtained. The component *A* and the compound *AB* are then partially miscible in the solid state, and so also are *AB* and the component *B*. There is little fundamental difference between this type of system and that for compound formation without miscibility of the solids (Fig. 179); in each case the maximum of the curve represents the composition and melting point of the compound.

Changes in the Solid State.—Solid phases frequently undergo transformations; not only may the compositions of solid solutions alter, but the changes may be more fundamental. One change of special interest arises when, on

cooling a solid solution, crystals of one or other of the components "separate." This behavior is exactly comparable with the deposition of pure solid from a liquid solution, and in fact the curves representing equilibrium between pure solid and solid solution are similar to those for solid and liquid solution. The simplest type of behavior is analogous to the formation of a simple eutectic; this is illustrated in Fig. 190, the upper curves representing the equilibrium between liquid and a continuous series of solid solutions. As the solid solution *S* is cooled it begins to separate pure *A* along *CD*, or pure *B* along *ED*. At the point *D* the solid solution *S* is in equilibrium with solid *A* and solid *B*, so that the system is invariant; *D* is in fact analogous to a eutectic, except that a solid solution replaces the liquid. The point *D* is called a **eutectoid** (H. M. Howe, 1903). The properties of the system represented by *CDE* can be deduced completely by comparison with Fig. 171. It is not necessary that the components *A* and *B* should form a continuous series of solid solutions; any single solid solution is theoretically capable of forming a eutectoid. A eutectoid point has been observed in the iron-carbon system; the solid solution *austenite* separates out either pure iron, on one side of the eutectoid, or iron carbide (*cementite*) on the other. The eutectoid temperature is 721° C. and the system contains 0.9 per cent. of carbon by weight.

Changes in the solid state are of importance in metallurgical work, and various kinds are known in addition to that described above. As an illustration it may be mentioned that copper and gold form a complete series of solid solutions with a minimum melting point, but if the solids are cooled slowly Cu_3Au and CuAu are formed (cf. p. 381).

Systems Involving a Salt and Water.—When the two components are a salt and water, it is possible to obtain the complete equilibrium diagram in only a small number of cases by working at high pressures. If observations are made at atmospheric pressure there is a limit to the concentration range which can be studied; apart from this factor there is, however, no fundamental difference between the properties of salt-water systems and others of two components, although it may be convenient to consider them from a somewhat different point of view.

In Fig. 191 is seen the equilibrium diagram for potassium iodide and water at atmospheric pressure; no compounds are formed and the system is of the simple eutectic type. The curves, as in the other diagrams, represent the composition of the liquid phase in equilibrium with either solid water, i.e., ice, or solid potassium iodide at various temperatures. Along *AC* ice separates, and it is convenient to regard this curve as giving the *freezing points* of solutions of potassium iodide in water; along *BC* the solid deposited is potassium iodide, and so it may be considered as the *solubility curve* for that salt. These names indicate the methods whereby the two curves are obtained, viz., by freezing-point and solubility measurements, respectively. The steep rise of the curve *CB* shows that the solubility of potassium iodide in water increases relatively slowly with temperature; other substances show a more marked variation, indicating that they have a greater heat of solution. The freezing-point and solubility curves meet at *C*, the eutectic point, where both ice and potassium iodide separate at -23° , this being the lowest temperature at which the

solution can exist. All solutions when cooled will ultimately show a halt in temperature at this point until the liquid phase has disappeared. Eutectic systems involving a salt and water were the first to be studied (F. Rüdorff, 1864), and to them F. Guthrie (1875) gave the name **cryohydrates** (Greek: *frost water*). From the fact that systems having the cryohydrate composition melt and freeze sharply at a definite temperature, the **cryohydric point**, Guthrie concluded that they were definite compounds, but the work of L. Pfaundler (1877) and of H. Offer (1880) showed this to be incorrect. The physical properties of the cryohydrates, e.g., heat of solution and density, are the mean of the values for the two components, suggesting that they are mixtures, and their heterogeneous nature can be observed under the microscope, when the separate crystals

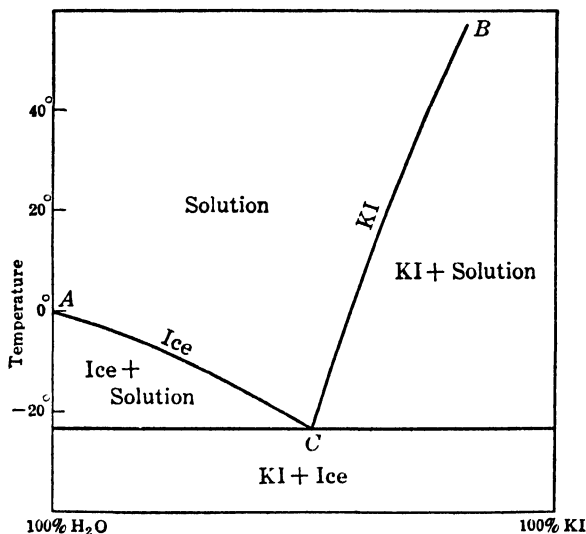


FIG. 191. System of a salt (KI) and water

of ice and of the salt can be identified. The most important argument, however, is that the phase rule requires there should be two solid phases, and not one, at the eutectic (cryohydric) point. The terms cryohydrate and cryohydric point are obsolescent and will not be employed further. The possibility that the salt may exist in several crystalline forms, e.g., ammonium nitrate, leading to discontinuities in the solubility curve has been mentioned previously (p. 752).

Freezing Mixtures.—The familiar production of a freezing mixture with a salt and ice is readily explicable by means of the equilibrium diagrams. If a salt is added to ice and a little water at 0° c., some of the salt will dissolve in the water, so that there will now be salt, ice and solution. Such a system can only be in stable equilibrium at the eutectic

point, which is much lower, and hence ice will melt and salt will continue to dissolve in the water produced. Melting of the ice is always, and dissolution of the salt is generally, accompanied by absorption of heat, and so the temperature of the system will fall; this fall will continue until one of the solid phases is used up. If there is a relatively large quantity of ice and salt present, however, the temperature must decrease until the eutectic point is reached. The mixture will stay at this temperature until sufficient heat is absorbed from the surroundings to cause all the ice to melt or all the salt to dissolve; there being now only one solid phase, the condensed system is univariant and the temperature will rise.

Hydrates with Incongruent Melting Points.—Since the formation of salt-water compounds with incongruent melting points, usually called transition points for this type of system, are relatively common, these will be considered first. A case which has been well studied since J. L. Gay-Lussac (1819) first observed the unusual behavior, is the system sodium sulfate-water, the phase diagram being depicted schematically in Fig. 192. Along curve *AB* the liquid phase is in equilibrium with ice,

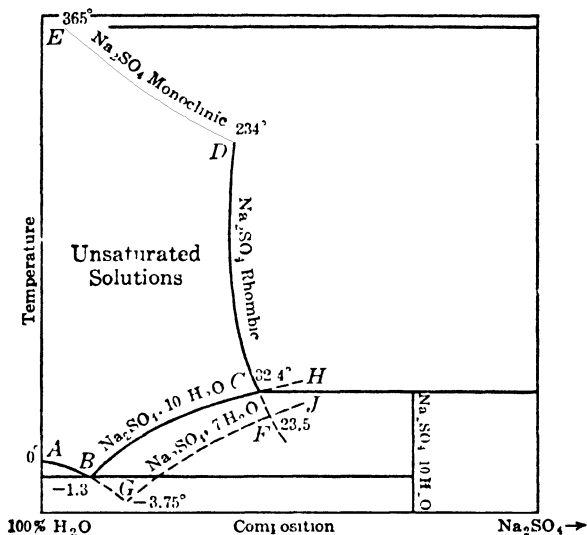


FIG. 192. Salt hydrates with incongruent melting points (Na_2SO_4 - H_2O)

whereas along *BC* the solid phase is $\text{Na}_2\text{SO}_4 \cdot 10\text{H}_2\text{O}$; the latter may thus be regarded as the solubility curve of the decahydrate. Before the curve *BC* reaches a maximum, however, an incongruent melting point, or transition point, is reached at *C*, when rhombic crystals of anhydrous sulfate commence to separate. At *C* the decahydrate and rhombic anhydrous salts are in equilibrium with saturated solution, and the condensed system is invariant. This transition temperature is so definite

at atmospheric pressure, viz., 32.383°C ., that it has been suggested as a fixed point in thermometry. The curve CD gives the compositions of solutions in equilibrium with rhombic anhydrous salt, and is the solubility curve of this substance. It will be noticed that up to about 125°C . the amount dissolved decreases with increasing temperature, but between 125° and 234°C . the solubility increases with temperature. At D , i.e., 234°C ., a further transition occurs between the rhombic anhydrous and monoclinic anhydrous sodium sulfate, and here both anhydrous forms are in equilibrium with saturated solution. At still higher temperatures the monoclinic salt only is stable in contact with solution, and it has a very marked negative temperature coefficient of solubility, as indicated by the curve DE . Since a saturated solution of sodium sulfate boils at about 103.3° under atmospheric pressure, all measurements above this temperature must be made at higher pressures. Even increase of pressure, however, sets a limit to the range over which the system can be studied, for the point E , i.e., 365°C ., is the critical point of water in the solution, and above this temperature solutions of sodium sulfate in liquid water cannot exist. The temperature 365°C . is actually quite close to the critical point of pure water, because the amount of sodium sulfate dissolved is then so small.

If a solution saturated with respect to anhydrous rhombic sodium sulfate is cooled along DC , then when the transition point C is reached the decahydrate should commence to deposit; sometimes, however, this does not happen, the anhydrous form remains metastable and the composition of the solution continues along DC beyond C until F is reached, at 23.5°C .,* when the heptahydrate $\text{Na}_2\text{SO}_4 \cdot 7\text{H}_2\text{O}$ separates. This substance is always metastable in the presence of the decahydrate, and it can only be formed if $\text{Na}_2\text{SO}_4 \cdot 10\text{H}_2\text{O}$ is completely absent. In these circumstances the metastable curves BG and GF , the latter being the solubility curve of the heptahydrate, can be realized, G being a metastable eutectic. The point F is the transition point at which heptahydrate and anhydrous salt are in metastable equilibrium with solution. The fact that the metastable solubility curves GF and FC are found to the right of the stable curve BC , shows that at a given temperature the metastable forms are more soluble than a stable form. The addition of a small crystal of decahydrate will cause the excess salt to be precipitated in that form, and the composition will fall on to the line BC . The statement that a metastable form is more soluble than the stable form at the same temperature represents a rule of universal applicability, which can be derived from thermodynamic considerations. The metastable form can be converted into a stable form as a result of a spontaneous change, but the reverse never occurs; hence the stable form has the lower chemical potential. Since this potential is the same for the solute in a saturated solution as for the solid with which it is in equilibrium, it follows that the chemical potential of the metastable solute in its saturated solution is greater than that of the stable form in a saturated solution at the same temperature.

* The exact value is given as $23.465 \pm 0.004^{\circ}\text{C}$.

If the two solutions behave ideally, or depart from ideality to the same extent, as is very probable, the former will thus be more concentrated than the latter.

In addition to the curves BG , GF and FC , it is possible, on account of suspended transformation, for the solubility curve BC of the decahydrate to be continued into the metastable region CH , and that of the heptahydrate along FJ . In each case the metastable solutions are supersaturated with respect to the stable solutions at the same temperature. The dependence of solubility on the nature of the solid phase in contact with the solution is evident, and hence in stating solubility data it is important always to define the exact nature of the salt with which the solution is saturated.

In the system sodium chloride and water the eutectic temperature is -21.2°C ,* the solid phases being ice and $\text{NaCl}\cdot 2\text{H}_2\text{O}$; the hydrate is, however, unstable above 0.15°C , and at this temperature it is in equilibrium with the anhydrous salt and saturated solution. This temperature is, in fact, an incongruent melting point, and is the transition point for the system $\text{NaCl}\cdot 2\text{H}_2\text{O} \rightleftharpoons \text{NaCl} + \text{satd. solution}$. Another system having hydrates with incongruent

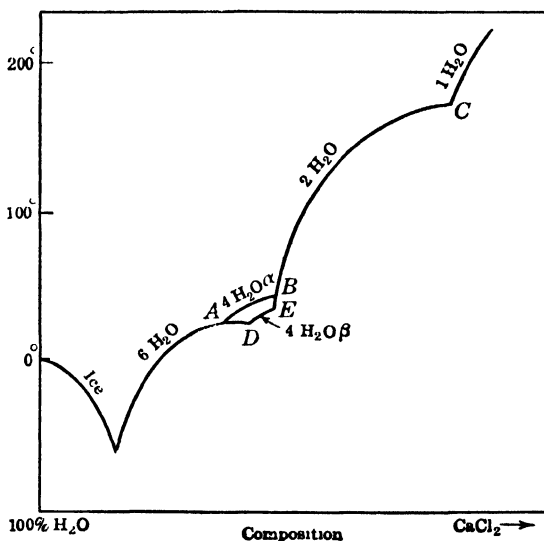
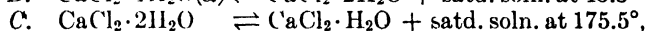
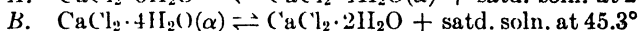
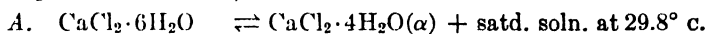


FIG. 193. Solubility curves of hydrates of CaCl_2

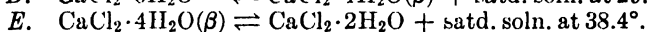
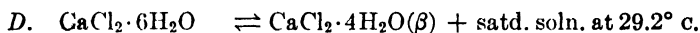
melting points is calcium chloride-water, studied by H. W. B. Roozeboom (1889); four stable hydrates, viz., $\text{CaCl}_2\cdot 6\text{H}_2\text{O}$, $\text{CaCl}_2\cdot 4\text{H}_2\text{O}(\alpha)$, $\text{CaCl}_2\cdot 2\text{H}_2\text{O}$ and $\text{CaCl}_2\cdot \text{H}_2\text{O}$ have been identified. In addition a metastable hydrate $\text{CaCl}_2\cdot 4\text{H}_2\text{O}(\beta)$ has been found to exist; its solubility curve is shown at DE below that for the stable tetrahydrate (Fig. 193). There are three stable

* When ordinary salt and water are mixed the eutectic is generally the metastable one for the anhydrous sodium chloride and ice, viz., -22.4° .

transition points shown, viz.,



and one outside the diagram for $\text{CaCl}_2 \cdot \text{H}_2\text{O} \rightleftharpoons \text{CaCl}_2 + \text{satd. solution}$, which is only known approximately as about 260°C. In addition there are two metastable transition points, viz.,



It will be noted that the hexahydrate and the dihydrate become unstable at temperatures very close to their respective melting points; if the conditions are such that the metastable (β) tetrahydrate separates, the hexahydrate has actually a congruent melting point.

Influence of Impurities on Transition Point.—The presence of impurities in the salt hydrate almost invariably depresses the transition point: the system now becomes one of three components, and hence the transition point is no longer a definite temperature, at a fixed pressure, but its value depends on the amount of the added substance. For relatively small amounts of the latter the depression of the transition point is approximately proportional to its molar concentration, and attempts have been made to utilize this fact for the determination of approximate molecular weights.

Determination of Transition Points.—The general procedures for the measurement of transition points have been considered in Chapter VI, and the methods described there can all be adapted to the study of salt hydrates (see also p. 464).

Hydrates with Congruent Melting Points.—One of the best known systems involving salt hydrates with congruent melting points is that formed by ferric chloride and water (H. W. B. Roozeboom, 1892); four stable hydrates, each having a congruent melting point, have been found. These are generally given the formulae $\text{Fe}_2\text{Cl}_6 \cdot 12\text{H}_2\text{O}$, $\text{Fe}_2\text{Cl}_6 \cdot 7\text{H}_2\text{O}$, $\text{Fe}_2\text{Cl}_6 \cdot 5\text{H}_2\text{O}$ and $\text{Fe}_2\text{Cl}_6 \cdot 4\text{H}_2\text{O}$, the double formula being used for ferric chloride in order to avoid a fractional number of molecules of water of crystallization. It will be evident from Fig. 194 that each hydrate has a definite melting point; it may alternatively be regarded, in each case, as the temperature at which the saturated solution of the hydrate has the same composition as the solid phase. In addition to the five stable eutectics there is a metastable one between $\text{Fe}_2\text{Cl}_6 \cdot 12\text{H}_2\text{O}$ and $\text{Fe}_2\text{Cl}_6 \cdot 5\text{H}_2\text{O}$, which is not shown; this point can be attained because the hydrate $\text{Fe}_2\text{Cl}_6 \cdot 7\text{H}_2\text{O}$ is somewhat reluctant to crystallize, and so it is possible to continue the curves for the other two hydrates until they meet. The solutions are, of course, supersaturated with reference to the stable hydrate at the same temperature.

If it is required to determine the behavior of any particular unsaturated solution on cooling, the method of treatment is identical with that already

described in connection with two-component systems in general; the fact that one of the components is water makes no difference to the nature of the results. The subject need, therefore, not be considered further. It is of interest, however, to enquire into the behavior on evaporation at constant temperature, and the system ferric chloride-water may be used for purposes of illustration. The phenomena observed when solutions of ferric chloride are evaporated are somewhat remarkable, but they can be readily explained with the aid of the equilibrium diagram; an enlarged, slightly exaggerated, portion

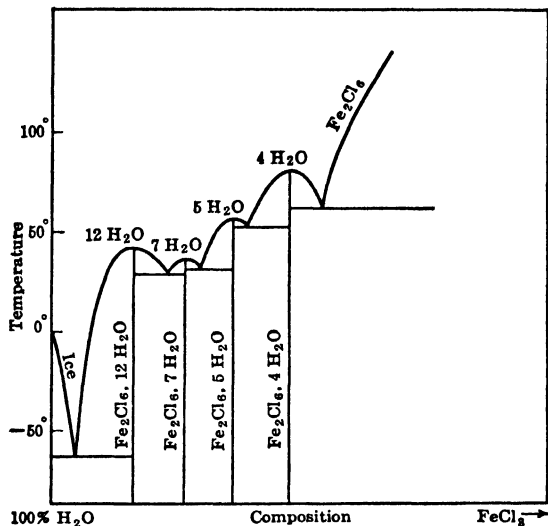


FIG. 194. Salt hydrates with congruent melting points ($\text{FeCl}_3\text{-H}_2\text{O}$)

of this is shown in Fig. 195 in order to emphasize the salient points. Suppose an unsaturated solution *a* is evaporated at constant temperature; the composition of the system as a whole will move along the horizontal line *a* to *l*. When the point *b* is reached, $\text{Fe}_2\text{Cl}_6 \cdot 12\text{H}_2\text{O}$ begins to separate, the solution having become saturated with respect to this salt. As evaporation proceeds between *b* and *c*, the composition of the liquid remains at *b* but its volume becomes less and less while the amount of solid hydrate increases. The condensed system involving a salt and its saturated solution has one degree of freedom, apart from pressure, and if the temperature is fixed the composition of the liquid phase cannot alter. At *c* the composition of the system is $\text{Fe}_2\text{Cl}_6 \cdot 12\text{H}_2\text{O}$, and since the temperature is below the melting point of the compound, it is evident that at this point the liquid phase has entirely disappeared, and the solid hydrate alone remains. As evaporation proceeds beyond *c*, however, the solid commences to liquefy, the liquid phase having the composition *d*; this continues until the point *d* is reached when the whole system is once more liquid. It remains in this state between *d* and *e*, which is a region of unsaturated solutions, and at *e* a new hydrate $\text{Fe}_2\text{Cl}_6 \cdot 7\text{H}_2\text{O}$ commences to separate. Again the quantity of liquid, the composition of which remains at *e*, decreases while the

solid increases, until the point *f* is reached when the whole consists of solid $\text{Fe}_2\text{Cl}_6 \cdot 7\text{H}_2\text{O}$. Between *f* and *g* there is further liquefaction, *g* being the composition of the liquid phase, and when the whole system has the composition *g* it is once more completely liquid. Between *g* and *h* the solution is again unsaturated, but once more a solid hydrate $\text{Fe}_2\text{Cl}_6 \cdot 5\text{H}_2\text{O}$ forms at *h*, and the

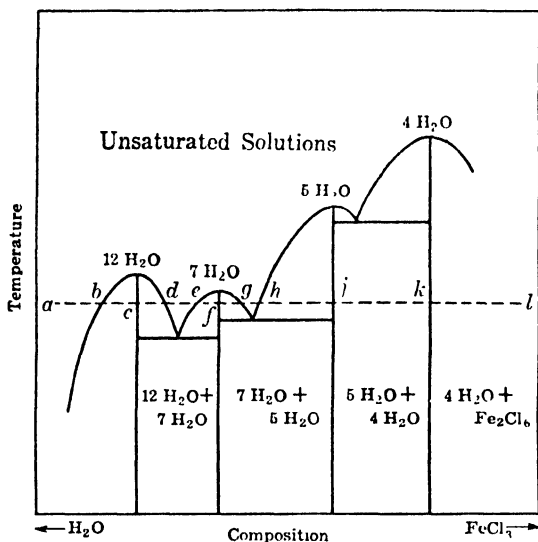


FIG. 195. Isothermal evaporation of aqueous solutions of ferric chloride

whole system becomes solid at *j*. Further dehydration cannot result in further liquefaction, since the temperature is now below that at which unsaturated solutions exist, but at *j* the next lower hydrate $\text{Fe}_2\text{Cl}_6 \cdot 4\text{H}_2\text{O}$ commences to form; between *j* and *k* the penta- and tetra-hydrates are in equilibrium, and at *k* the latter only will remain. Continued dehydration leads to the gradual formation of the anhydrous salt which will eventually be the only residue.²³

It may be mentioned here, although the components are not salts, that a number of hydrates, all having congruent melting points, have been discovered in the systems $\text{SO}_3\text{-H}_2\text{O}$ and $\text{N}_2\text{O}_5\text{-H}_2\text{O}$ respectively: they are $\text{SO}_3 \cdot 5\text{H}_2\text{O}$ (-25°), $\text{SO}_3 \cdot 3\text{H}_2\text{O}$ (-39°), $\text{SO}_3 \cdot 2\text{H}_2\text{O}$ (8.5°), $\text{SO}_3 \cdot \text{H}_2\text{O}$ (10.4°), and $2\text{SO}_3 \cdot \text{H}_2\text{O}$ (36°); and $\text{N}_2\text{O}_5 \cdot 7\text{H}_2\text{O}$ (-18.5°), $\text{N}_2\text{O}_5 \cdot 3\text{H}_2\text{O}$ (-38°), and $\text{N}_2\text{O}_5 \cdot \text{H}_2\text{O}$ (-41.2°). The temperatures in parentheses are the respective melting points.

GAS-SOLID SYSTEMS

The nature of the equilibrium between a gas and a solid depends on whether the two combine to form one or more compounds, or whether the

gas dissolves in the solid, with or without compound formation. The most important type actually observed involves compound formation but not solution, and this will be considered first.

Compound Formation between Gas and Solid.—A simple example of gas-solid equilibrium with compound formation is that between solid calcium oxide and gaseous carbon dioxide, the product being solid calcium carbonate. This system has two components (p. 475), and when the gas and the two solids, which are not soluble in each other, are in equilibrium there are three phases; hence by the phase rule

$$F = C - P + 2 = 2 - 3 + 2 = 1.$$

The system is consequently univariant, and this means that only one variable, e.g., temperature, need be fixed to define the system completely; the pressure of carbon dioxide gas in equilibrium with solid calcium oxide and carbonate should thus be a constant at each definite temperature, independent of the amounts of the various reacting substances present. By application of the law of equilibrium an identical conclusion is reached, and this has been verified experimentally (p. 845). The equilibrium pressure at each temperature is known as the **dissociation pressure**; it varies with temperature in the same manner as the vapor pressure of a pure substance, and in fact the simplified form of the Clapeyron-Clausius equation (p. 452) is applicable. It is more customary, however, to consider this aspect of the subject in connection with the study of the law of mass action, and so further discussion will be postponed.

If the pressure of carbon dioxide, over calcium oxide and carbonate, is maintained below the dissociation pressure, the carbonate will dissociate completely, whereas if the pressure is kept above the equilibrium value the calcium oxide will continue to react with the carbon dioxide either until the former is all used up, or until the pressure of the gas is reduced. It is now possible to understand what happens when carbon dioxide is steadily admitted into a closed space containing calcium oxide at constant temperature. No reaction will occur until the pressure of gas has risen to the dissociation pressure at that temperature; then combination takes place and the pressure remains constant until all the calcium oxide has been converted to carbonate, when the pressure will rise again (Fig. 196). The same phenomena occur in exactly the reverse order if the pressure is gradually decreased to a very small value.

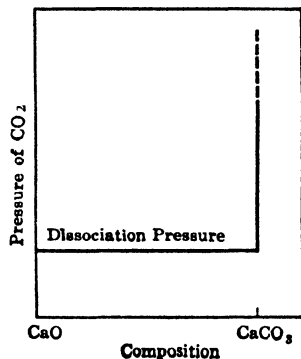
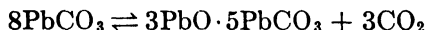


FIG. 196. Dissociation of calcium carbonate

When a metallic oxide combines with carbon dioxide to form more than one compound, there will be more than one constant pressure stage;

this occurs with the oxides of magnesium and lead, a number of basic carbonates being possible, especially in the latter case. The pressure of carbon dioxide, at a definite temperature, should increase until the dissociation pressure of the first basic carbonate is reached, and then remain constant until the formation of this compound is complete; then the gas pressure should rise to that of the dissociation pressure of the next basic carbonate, and so on. In the actual investigation it was found more convenient to keep the pressure of carbon dioxide constant, and alter the temperature so as to bring about dissociation. Starting with pure lead carbonate and raising the temperature gradually, it was observed that under a carbon dioxide pressure of 1 atm. no decomposition occurred until a temperature of 274° c. was attained; the product of dissociation is then the basic carbonate $3\text{PbO} \cdot 5\text{PbCO}_3$. The dissociation pressure of the system



is thus 1 atm. at 274° c. In agreement with the requirements of the phase rule, the temperature remained constant as long as the two solids were present; when, however, the normal carbonate had been completely decomposed into the basic salt, a further increase of temperature was necessary, to 286° c., before dissociation could be continued, and another basic compound $\text{PbO} \cdot \text{PbCO}_3$ was formed. Two further increases of temperature, at 360° and 412° c., were also noted, the new compounds produced being $2\text{PbO} \cdot \text{PbCO}_3$ and PbO , respectively. When lead carbonate is heated and the pressure of carbon dioxide maintained at 1 atm., the temperature must be raised to 412° c. before decomposition can be complete. At a lower carbon dioxide pressure a lower temperature would, of course, be adequate for the purpose. If the carbon dioxide is removed continuously, complete dissociation can occur at relatively low temperatures, but by making the observations as described above it is possible, by occasional analysis of the solid phase, to study all the intermediate stages in a dissociation reaction.

The procedure just described has also been used to study the decomposition of metallic oxides; by heating the oxide in an open vessel the pressure of oxygen is kept at a constant value of 0.21 atm., its partial pressure in the atmosphere. The temperatures at which various stages of dissociation occur can then be determined, the nature of the products being found by analysis, or by the loss in weight. The study of the dissociation of metallic oxides is often rendered complicated by the fact that when a higher oxide is heated the lower or intermediate oxide may be produced which does not form a separate phase but exists as a solid solution in the higher oxide. In such a case there is only one solid phase, and hence the system is bivariant. If the pressure of oxygen is fixed, therefore, the temperature will not remain constant but will change steadily as the composition of the solid solution alters. Similarly, if the temperature is kept constant the dissociation pressure will change con-

tinuously. If these phenomena are observed then it must be concluded that the substances present in the system have formed solid solutions with one another.

Salt Hydrates and Ammines.—As shown above, two methods are available for the study of gas-solid equilibria: either the temperature is kept constant and the pressure determined as the composition of the solid phase changes, or the pressure is held constant and the temperature necessary to attain that dissociation pressure is observed. The former is most commonly used in the study of salt hydrates; these constitute gas-solid systems, the gas phase being water vapor, and their behavior is similar to that of the systems already considered. Starting with the anhydrous salt it is possible to admit water vapor gradually and to measure the equilibrium pressure at constant temperature. Alternatively, the highest known hydrate may be dehydrated gradually by pumping off the water vapor, and the dissociation pressure determined from time to time. Another possibility is to obtain a series of mixtures of salt hydrates of known composition by any convenient method, e.g., heating the highest hydrate in air, and to measure the equilibrium pressure.

Experimental Methods.—The dissociation pressure of a salt-hydrate system is, of course, the pressure of water vapor; for this reason it is generally referred to as the vapor pressure of the system. The methods used for determining these pressures resemble those for the study of aqueous solutions. The first measurements were made by introducing the salt hydrate into the top of a barometer tube and observing the depression, but the results were very approximate. The direct method of P. C. F. Frowein (1887), described on p. 629, gave the first data of any value, and the tensimeter he designed has been improved and employed by a number of subsequent workers; precautions have to be taken to insure the almost complete elimination of air from the apparatus. The dew-point method (p. 632) has also been used for the study of salt hydrates. The best results seem to have been obtained by the transpiration method, first applied to the study of salt hydrates by G. Tammann (1888) and improved by a number of different investigators. The procedure is very similar to that already considered (p. 448). Some workers prefer to measure the volume of air passing through the apparatus, whereas others do not do so, but draw the same quantity of air through the salt hydrate system and through pure water, the vapor pressure of the latter being known. It will be noted that the use of pre-saturators is unnecessary, since the vapor pressure remains constant provided the higher hydrate is not completely decomposed. A form of isopiestic method (p. 632) was devised by W. Muller-Erzbach (1881); the concentration of aqueous sulfuric acid was found over which the salt hydrate system neither gained nor lost weight. The vapor pressures must then be equal, and the value can be obtained from tabulated data on mixtures of water and sulfuric acid. A different principle was first applied by C. E. Linebarger (1894) and modified later (R. E. Wilson, 1921). The salt hydrate system is shaken with an organic liquid, such as ether or isoamyl alcohol, in which the salts are insoluble but water is slightly soluble. When equilibrium is reached the vapor pressure of the water dissolved in the liquid is equal to that of the salt hydrate system; the proportion of water is found, e.g., by adding a known

amount of a suitable electrolyte and determining the conductance, and the vapor pressure found from data obtained previously by measurements on the organic liquid-water system. On account of the possibility of suspended transformation, it is desirable in all methods for measuring vapor pressure that the salt should be partially dehydrated, e.g., by heating, prior to its use in the experiment, so as to ensure the presence of both solid phases²⁴

Results.—The general nature of the results obtained may be illustrated by consideration of the hydrates of copper sulfate at 25° c.; starting with $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ and dehydrating gradually, the equilibrium pressure of water vapor is found to be 7.8 mm. of mercury, and to remain constant as long as the system contains more water than $\text{CuSO}_4 \cdot 3\text{H}_2\text{O}$; when this point is reached the pressure falls sharply to 5.6 mm. It then stays constant at this value until the composition of the system is $\text{CuSO}_4 \cdot \text{H}_2\text{O}$, when it again falls to 0.8 mm. at which it stays until only anhydrous salt remains, and the vapor pressure of water falls to zero. It

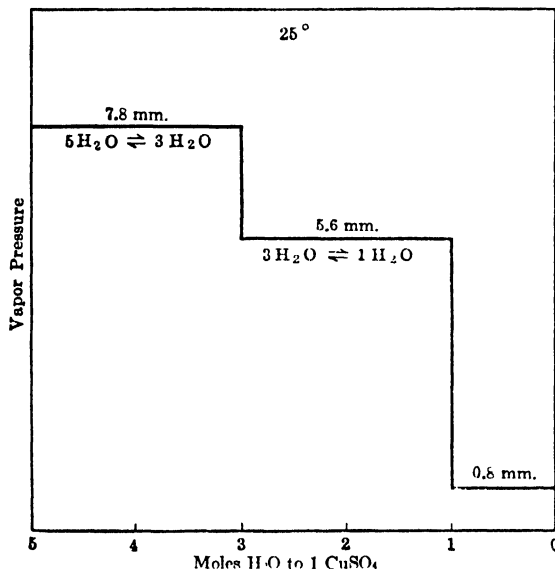
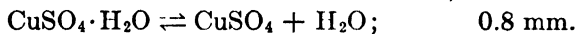
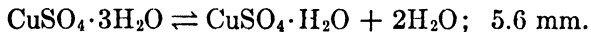
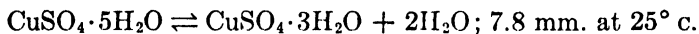


FIG. 197. Dehydration of copper sulfate at 25°

is evident from these results, which are plotted in Fig. 197, that three equilibria are to be considered, viz.,



If the starting material had been anhydrous copper sulfate, to which water was added gradually, the same results would have been observed in the

reverse direction. The dehydration of the copper sulfate hydrates has been studied at other temperatures, and similar results obtained, except that the vapor pressures are higher the higher the temperature; at 50° c., for example, the three systems mentioned above give equilibrium vapor pressures of 45.4, 30.9 and 4.5 mm. respectively. It appears that above about 102° the pentahydrate is not stable, and $\text{CuSO}_4 \cdot 3\text{H}_2\text{O}$ is the first stable hydrate; at such temperatures only two horizontal portions would be found in the vapor pressure-dehydration diagram.*

It is important to emphasize that the existence of a definite vapor pressure of a salt-hydrate system at a particular temperature requires the presence of *two solid phases* in addition to vapor, for only then is the system univariant. It is not correct, therefore, to speak of the "vapor pressure of a salt hydrate," for the hydrate itself has no definite water vapor pressure, as may be seen from Fig. 197; both an upper and lower hydrate, or anhydrous salt, must be present simultaneously. The vapor pressure for the *system* is then definite and independent of the relative amounts of the two hydrates. Further, the actual vapor pressure will depend on the nature of the hydrates in equilibrium; for example, although the system $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ - $\text{CuSO}_4 \cdot 3\text{H}_2\text{O}$ has a vapor pressure of 7.8 mm. at 25° c., if by some means it were possible to obtain the system $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ - CuSO_4 anhydrous, without the intervention of the intermediate hydrates, the vapor pressure would be 4.3 mm.

If water vapor at these low pressures behaves as an ideal gas, then the free energy change accompanying the transfer of 1 mole of water vapor from pure water to a particular hydrate is given by (see p. 855),

$$-\Delta F = RT \ln p/p_{A,B}, \quad (84)$$

where p is the vapor pressure of water, and $p_{A,B}$ that of the system resulting when water is removed from the hydrate A to form hydrate B, at the same temperature T . The change of free energy resulting from the transfer of 5 moles of water from pure water to CuSO_4 anhydrous to form $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ in one stage, must be the same as that accompanying the transfer in three stages, viz., CuSO_4 to $\text{CuSO}_4 \cdot \text{H}_2\text{O}$, then $\text{CuSO}_4 \cdot \text{H}_2\text{O}$ to $\text{CuSO}_4 \cdot 3\text{H}_2\text{O}$, and finally $\text{CuSO}_4 \cdot 3\text{H}_2\text{O}$ to $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$; thus,

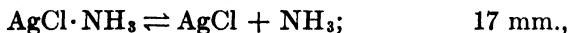
$$5RT \ln \frac{p}{p_{5,0}} = RT \ln \frac{p}{p_{1,0}} + 2RT \ln \frac{p}{p_{3,1}} + 2RT \ln \frac{p}{p_{5,3}}, \quad (85)$$

$$\therefore p_{5,0}^5 = p_{1,0} \times p_{3,1}^2 \times p_{5,3}^2, \quad (86)$$

where the subscripts indicate the number of molecules of hydrate water in the two salts constituting the system. It is thus calculated that $p_{5,0}$ should be 4.3 mm. at 25°, as given above.

* From careful dehydration experiments T. I. Taylor and H. P. Klug (1936) have concluded that $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ changes to a *tetrahydrate* $\text{CuSO}_4 \cdot 4\text{H}_2\text{O}$ at 96.5°, and this is subsequently converted into $\text{CuSO}_4 \cdot 3\text{H}_2\text{O}$ at 102° and to $\text{CuSO}_4 \cdot \text{H}_2\text{O}$ at 115° c. They consider that previous workers have failed to detect $\text{CuSO}_4 \cdot 4\text{H}_2\text{O}$ because its vapor pressure is probably close to that of $\text{CuSO}_4 \cdot 3\text{H}_2\text{O}$.

The equilibria between a salt and ammonia gas are quite analogous to those involving a salt and water vapor; cases of this type, in which a number of metallic halides are concerned, have been studied by W. Biltz (1920-21). Silver chloride, for example, forms three compounds, viz., $\text{AgCl} \cdot 3\text{NH}_3$, $2\text{AgCl} \cdot 3\text{NH}_3$, and $\text{AgCl} \cdot \text{NH}_3$, so that the following three equilibria are observed:



the corresponding dissociation pressures being recorded. The change from one pressure to the next takes place sharply when the appropriate composition is reached. Above 33° c. the compound $\text{AgCl} \cdot \text{NH}_3$ dissolves to some extent in the solid $2\text{AgCl} \cdot 3\text{NH}_3$; hence, as already explained, the dissociation pressure over a range of composition does not remain constant but varies with the proportion of the components.

Influence of Temperature.—The dissociation pressure generally increases with increasing temperature because the chemical changes involved are almost invariably accompanied by an absorption of heat; from measurements of the pressures at two temperatures it is possible to calculate the heat of dissociation. When the system is one of salt hydrates this is equal and opposite to the heat of hydration of the lower hydrate with water vapor. This aspect of the subject will be considered more fully in Chapter XI, and a qualitative treatment of the influence of temperature in connection with salt hydrates will be considered below.

Solution of Gas in Solid.—In the cases already discussed the gas itself was assumed to be insoluble in the solid; if it dissolves, however, from the commencement, there will be two phases, viz., gas and solution of gas in solid. According to the phase rule, since $C = 2$ and $P = 2$, the system has two degrees of freedom; hence at constant temperature the composition of the solid phase, that is the amount of gas dissolved, will vary with the gas pressure. This may continue until the solid is saturated, when no more gas will dissolve and the pressure should increase sharply. There are, however, two other possibilities which may arise: either (i) the solution of the gas in the solid may, after a certain concentration is attained, break up into two solid solutions, or (ii) chemical combination may occur between the gas and solid when the dissociation pressure, at that temperature, is reached. One of the most interesting gas-solid systems in which solutions are formed is that involving palladium and hydrogen. Although it has been the subject of a great deal of investigation since T. Graham (1867) discovered that palladium "occludes" large volumes of hydrogen, the behavior of the system is still not completely understood.

The form of the curves representing the amount of hydrogen taken up by palladium at various pressures is shown in Fig. 198, and, although there may be some difference of opinion concerning details, the main features are

undisputed. As the pressure of hydrogen is gradually increased the quantity dissolved by the palladium also increases, since the system is bivariant; at low pressures the amount w taken up by a given weight of palladium is approximately proportional to the square root of the pressure p , i.e.,

$$w/\sqrt{p} = \text{constant (approx.)}.$$

Since in the gaseous state the hydrogen consists of H_2 molecules, this result implies, in accordance with the distribution law (p. 737), that in the palladium it exists in the atomic form; there are also other reasons for favoring this conclusion. After a certain pressure, which is smaller the lower the temperature, the pressure remains constant although occlusion of the hydrogen by the palladium continues, until the composition of the solid corresponds approximately to Pd_2H , when the pressure again rises. The horizontal portion of the curve should, by the phase rule, mean the

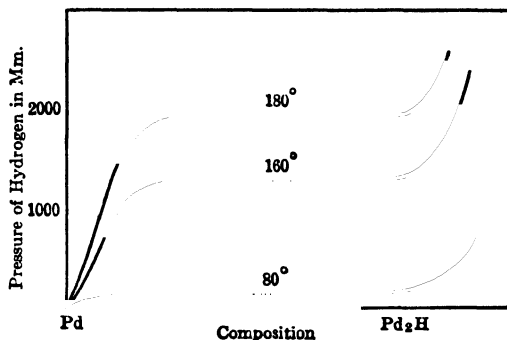


FIG. 198. The palladium-hydrogen system

formation of two solid phases, so that the system is now univariant, instead of bivariant. One possibility is that there are two immiscible solid solutions of hydrogen in palladium, but the fact that the pressure remains constant until the composition is Pd_2H suggests that a compound of this formula is produced. It is a very striking fact that many different workers, using a variety of samples of palladium, have always found the constant pressure region to end in the vicinity of the same composition, but it must be admitted that a compound of the formula Pd_2H is unusual. The explanation is that the so-called palladium hydride is not a true compound in the sense that covalent bonds are involved, but that it is an interstitial compound (p. 382). The hydrogen probably enters the palladium lattice in the atomic form and fills up spaces between the palladium atoms, and all the available sites are occupied when the ratio of palladium to hydrogen is 2 to 1. If this were the whole explanation then it would not be easy to understand why the pressure remains constant while entry into the lattice is occurring, for such a system would

constitute one solid phase only. It must be remembered, however, that surface phenomena play an important part in the palladium-hydrogen system, and that in such circumstances the phase rule no longer applies in its simple form. Another possibility is that the system does actually consist of two solid solutions, as indicated above; one of them would then be made up of Pd_2H in which further hydrogen gas can dissolve.²⁵

SOLID-LIQUID-VAPOR EQUILIBRIA

Salt Hydrates and Solutions.—In the study of liquid-solid equilibria for salt-water systems the vapor was ignored, and in connection with the examination of solid-gas equilibria the conditions were such that no liquid phase was present. It is necessary now to consider the solid-liquid-vapor equilibria for some systems involving a salt and water especially where hydrates are formed. In order to overcome the difficulty of representing three variables, viz., pressure, temperature and concentration, graphically the change of vapor pressure with temperature is recorded for all the possible states with three phases in equilibrium. For a two-component system, three phases mean that the number of degrees of freedom is unity, since

$$F = C - P + 2 = 2 - 3 + 2 = 1;$$

hence at every temperature the system will have a definite vapor pressure, and the data can be plotted on a two-dimensional diagram.

The first case to be considered will be sodium sulfate-water; the vapor pressure diagram is drawn, not to scale, in Fig. 199, the curves for metastable systems being shown dotted. Each of the curves, except those for pure water included for comparison, represents three phases in equilibrium, and in the areas between the curves the system consists of two phases. The region in which unsaturated solutions and vapor, and anhydrous salt and vapor exist are marked; in addition the area *BCM* is that in which the decahydrate and vapor are normally found. If the formation of this hydrate is prevented then the metastable heptahydrate and vapor occur in the area *GFL*. The curves for the systems $\text{Na}_2\text{SO}_4 \cdot 10\text{H}_2\text{O}$ -saturated solution-vapor; $\text{Na}_2\text{SO}_4 \cdot 10\text{H}_2\text{O}$ - Na_2SO_4 -vapor; and Na_2SO_4 -saturated solution-vapor meet at the point *C* (32.6° , E. Cohen, 1894) where the four phases, viz., decahydrate, anhydrous salt, saturated solution and vapor are in equilibrium. This is a **quadruple point**, and is a true invariant point for the system, since a two-component system of four phases has no degrees of freedom. It is the transition point between the decahydrate and anhydrous salt *at the equilibrium vapor pressure of the system* (30.8 mm. of mercury), and hence differs somewhat from the transition temperature measured at atmospheric pressure (p. 463). The vapor pressure curve for saturated solutions of the decahydrate meets the ice-vapor curve at the eutectic point *B*, which is also a quadruple point where decahydrate, saturated solution, ice and vapor are in equilibrium.

This is the eutectic point for the system measured at its equilibrium vapor pressure. The point *G* is similarly the corresponding metastable eutectic for heptahydrate, which can only be observed in the absence of decahydrate. It is important to note that at a given temperature the vapor pressure of the metastable system, $\text{Na}_2\text{SO}_4 \cdot 7\text{H}_2\text{O} \rightleftharpoons \text{Na}_2\text{SO}_4 + 7\text{H}_2\text{O}(g)$

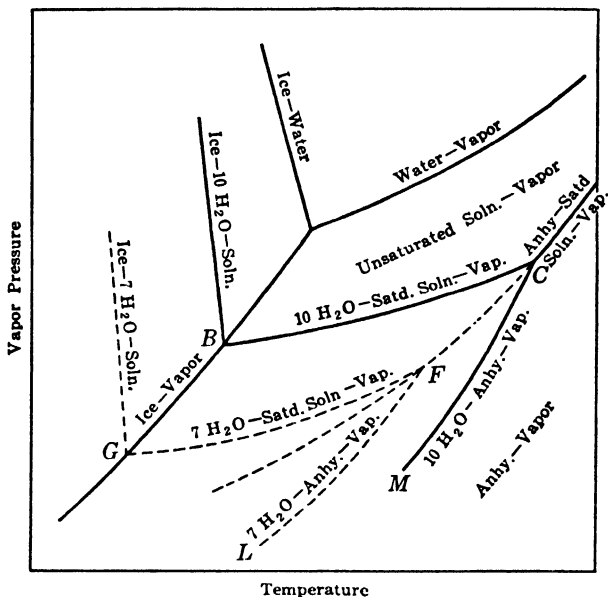


FIG. 199. Solid-liquid-vapor equilibria in the sodium sulfate-water system

(curve *LF*) is greater than that of the stable system $\text{Na}_2\text{SO}_4 \cdot 10\text{H}_2\text{O} \rightleftharpoons \text{Na}_2\text{SO}_4 + 10\text{H}_2\text{O}(g)$ (curve *MC*); this is in agreement with thermodynamic requirements that there should be a decrease of free energy in passing from a metastable to a stable state. On the other hand, the *saturated solution* of the metastable heptahydrate (curve *GF*) has a lower vapor pressure than that of the saturated solution of the stable decahydrate (curve *BC*), in agreement with the fact that the concentration of the former is greater than that of the latter (p. 774).

The vapor-solid equilibria for the copper sulfate-water system have been considered in some detail, and it is therefore of interest to see how the results may be correlated with the phase diagram on which liquid-phase equilibria are included. The complete data are not available, but Fig. 200, which is approximate only, is sufficient for the purpose; in view of the description given above there should be no difficulty in understanding the significance of the various curves. Suppose water is gradually admitted to the anhydrous salt while the system is kept at a constant temperature; starting from the point *a*, where the vapor pressure is zero,

the pressure-temperature conditions change along a vertical line from *a* to *f*. Between *a* and *b* the anhydrous salt and vapor exist without interaction; at *b* the monohydrate $\text{CuSO}_4 \cdot \text{H}_2\text{O}$ commences to form, and as long as this is in equilibrium with the anhydrous salt the vapor pressure remains constant, as required by the phase rule (p. 779). When the

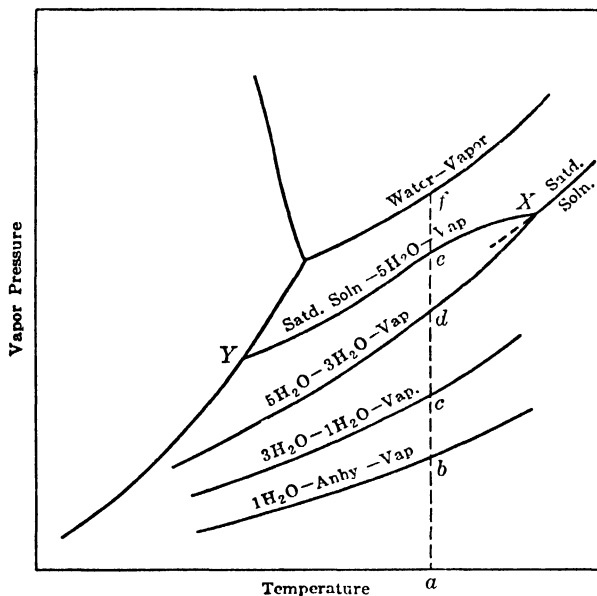


FIG. 200. Gradual hydration of copper sulfate

formation of $\text{CuSO}_4 \cdot \text{H}_2\text{O}$ is complete the further addition of a small amount of water results in the production of some $\text{CuSO}_4 \cdot 3\text{H}_2\text{O}$, so that the vapor pressure rises sharply to *c*. Here again the pressure remains constant until all the salt is in the form of the trihydrate. Once more, a further small quantity of water will be accompanied by formation of some $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ and a sudden rise of the pressure to *d*; the aqueous vapor pressure will then remain constant until the whole solid has been converted into the pentahydrate.* More water causes the latter to dissolve, forming a saturated solution, of vapor pressure *e*, and this will change only when the solid has dissolved completely. The vapor pressures from *e* to *f* are for a series of unsaturated solutions of increasing dilution.

If the process had been carried out above the transition point *X* of $\text{CuSO}_4 \cdot 5\text{H}_2\text{O} \rightleftharpoons \text{CuSO}_4 \cdot 3\text{H}_2\text{O} + \text{satd. solution}$, the trihydrate would dissolve without intervention of the pentahydrate, whereas below the eutectic point *Y* a mixture of a solid hydrate and ice would ultimately be obtained.

* The points *b*, *c* and *d* should be correlated with the three horizontal lines in Fig. 197.

Deliquescence and Efflorescence.—The results described above may be applied to account for the phenomenon of **deliquescence** (Latin: *to become liquid*). From what has been stated, it should be clear that, in general, if a salt is exposed to water vapor at a pressure greater than that of the saturated solution of its highest hydrate, at the given temperature, it will gradually take up water and finally form an unsaturated solution. A substance will be deliquescent, therefore, when the vapor pressure of its saturated solution is less than the usual pressure of water vapor in the atmosphere; water will continue to deposit until an unsaturated solution, having the same vapor pressure as water in the air, is formed. In order that the vapor pressure of the saturated solution may be sufficiently low, the salt should be very soluble, and hence it is only such substances, e.g., calcium chloride, a saturated solution of which has a vapor pressure of 4-5 mm. at room temperatures, that deliquesce. It will be obvious that deliquescence is a relative property, since it depends on the actual pressure of water vapor in the atmosphere which varies with time and place. In exceptionally dry atmospheres even calcium chloride will not deliquesce. Copper sulfate is not normally deliquescent, as at room temperatures the vapor pressure of its saturated aqueous solution is higher than the usual pressure of water vapor in the atmosphere; if the lower hydrates are exposed to normal air, however, the pentahydrate is formed. If the pressure of atmospheric water vapor is below 7.8 mm. at 25° c., as might possibly be the case, the hydration process would stop at the trihydrate.

The phenomenon of **efflorescence** (Latin: *to blossom*), i.e., the loss of water by a hydrated salt, is observed when the vapor pressure of a hydrate system is greater than that of water vapor in the air; dehydration will then occur in the effort for equilibrium to be attained between the hydrate system and its surroundings. Normally, the vapor pressures of the systems involving $\text{Na}_2\text{SO}_4 \cdot 10\text{H}_2\text{O}$, $\text{Na}_2\text{CO}_3 \cdot 10\text{H}_2\text{O}$ or $\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$ as the higher hydrate, are greater than that of atmospheric water vapor and so these substances are said to be efflorescent. Like deliquescence, the property of efflorescence is a relative one and depends on atmospheric conditions, for in a very dry atmosphere even $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ would be efflorescent, whereas in a very moist one sodium carbonate and sulfate might not possess this property.

Vapor Pressures of Saturated Solutions.—As the temperature is raised the amount of solute present at saturation generally increases, and so the normal effect of temperature in increasing the vapor pressure may be opposed by the diminution resulting from the higher concentration of the solution. If measurements could be carried out at sufficiently high pressures, and the salts did not decompose, the vapor pressures of saturated solutions would, in general, vary with temperature in the manner shown by the curve *ABC* in Fig. 201; discontinuities due to changes in the nature of the hydrates in equilibrium with the saturated solution have been ignored. The point *C* is the melting point of the anhydrous salt, where the vapor pressure is assumed to be zero. Since the

curve has a maximum, two possibilities may be expected to arise. If the maximum *B* occurs at a pressure below 1 atm. then the *saturated* solution will never boil under atmospheric pressure; this occurs with very soluble substances, such as sodium and potassium hydroxides. On the other hand, if the pressure at *B* is greater than atmospheric, then the saturated solutions can have two boiling points * as, for example, at *D* and *E*; solutions of potassium, sodium, silver and thallium nitrates, and sodium chlorate behave in this manner. The boiling points of saturated solutions of silver nitrate are at 133° and 191° c., the solid salt melting at 208° c.²⁶

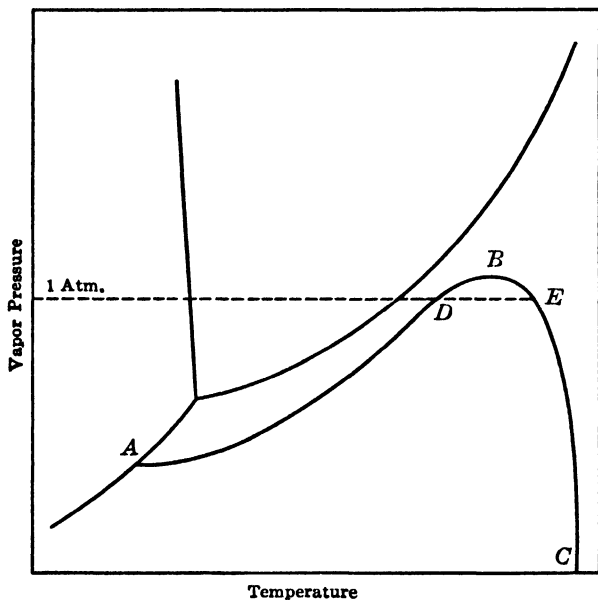


FIG. 201. Vapor pressures of saturated solutions

THREE-COMPONENT SYSTEMS

Graphical Representation.—For systems of three components, four degrees of freedom are possible, since temperature, pressure and the concentrations of two components may be independently variable. In order to simplify the graphical representation of the conditions of equilibrium for three components, the procedure generally adopted is to consider condensed systems, that is, the vapor is neglected, at constant temperature. By fixing two variables in this manner, only the two concentration terms can be independently varied and a planar diagram can be used to indicate equilibrium conditions. Although rectangular coordinates could be, and in some cases are, used (see p. 803), it is more customary with three-component systems to employ the triangular diagram, proposed by J.

* The concentrations of the two saturated solutions are, of course, different.

Willard Gibbs (1876), G. G. Stokes (1891) and W. H. B. Roozeboom (1894).

If ABC in Fig. 202 is an equilateral triangle, and P is any point within it, then the sum of the distances from P drawn parallel to the three sides is always the same, and is equal to the side of the triangle. By taking this side as unity, and expressing the amounts of the three components as fractions of the whole, it is possible to represent the composition of any ternary system by a point in the diagram. For example, in Fig. 202 the corners of the triangle may be assumed to represent the pure components A, B, and C; the distance from P to any side, measured parallel to either of the others, then gives the proportion of the component occupying

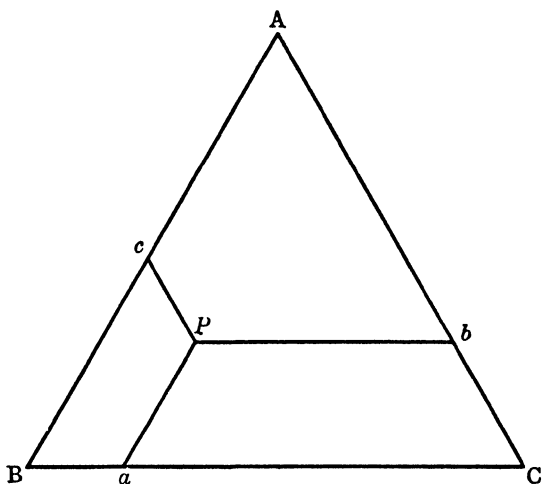


FIG. 202. Principle of the triangular diagram

the opposite corner; thus, the distances Pa , Pb and Pc give the fractional amounts of A, B and C respectively in the system indicated by the point P . Any point within the triangle represents three components, but a point situated on one of the sides indicates two components only; a point on the line BC would mean the amount of A is zero, and similarly along AC there is no B, whereas on the line AB the component C would be absent. In order to facilitate the plotting of compositions on the triangular diagram, the sides are divided into ten or more equal parts and a series of lines are drawn parallel to the three sides.

There are a number of properties connected with the equilateral triangle which make it particularly useful in connection with phase-rule studies. For example, if a line is drawn through any corner, e.g., A, to a point on the opposite side, e.g., to X (Fig. 203), then all points on the line AX represent a constant ratio of B to C, with variable amounts of A. Further, a line parallel to one of the sides, such as YZ , represents a con-

stant proportion of one component, in this case A, with variable amounts of the others. If P and Q indicate the compositions of any two mixtures of the three components, then any point, such as R , on the line joining them represents a mixture of P and Q , the amounts being in the proportion of RQ to RP , respectively. This property is useful if the composition of one phase, e.g., P , is known, and that of a mixture, e.g., R , of this and another phase Q ; the composition of the latter can then be ascertained by extrapolation.

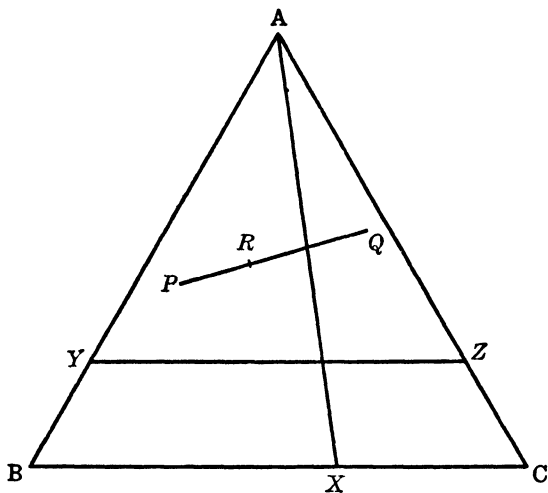


FIG. 203. Properties of the triangular diagram

In the study of three-component systems considerable complexity is possible; apart from vapor which is ignored, there may be three liquid and three solid phases, although according to the phase rule not more than four of these could be in equilibrium at any fixed temperature and pressure. In order to simplify the treatment, it will be assumed in the first place that no solid phases are present.

Systems of Three Liquids.—With systems of three liquids there are three main possibilities which may arise depending on the nature of the substances and the experimental temperature; they will be considered in turn.

(i) *One Pair of Partially Miscible Liquids.*—If the components are A, B and C, then it may be supposed that, at the given temperature, the liquids A and B are completely miscible, and so also are A and C, but B and C are only partially miscible. Examples of this type of system are acetic acid, chloroform and water; ethyl alcohol, ethyl acetate and water; and alcohol, benzene and water. Suppose the two partially miscible liquids B and C are taken; then at equilibrium there will be two conjugate layers whose compositions may be represented by the points b and c in

Fig. 204, since no A is present. Now imagine some A added to the system; as it is completely miscible with both B and C, it will distribute itself between the two layers which now become conjugate *ternary* solutions. The compositions of the two solutions will be given by points within the triangle, e.g., b' and c' ; these may be joined by a tie-line, the extremes of which give the compositions of two phases in equilibrium. It will be seen that the tie-line $b'c'$ slopes upwards to the right, indicating that the component A is relatively more soluble in the layer rich in C than it is in the one rich in B. This would be the case, for example, if A were acetic acid, and B and C were chloroform and water, respectively. Additions of further amounts of A not only cause the layers to dissolve

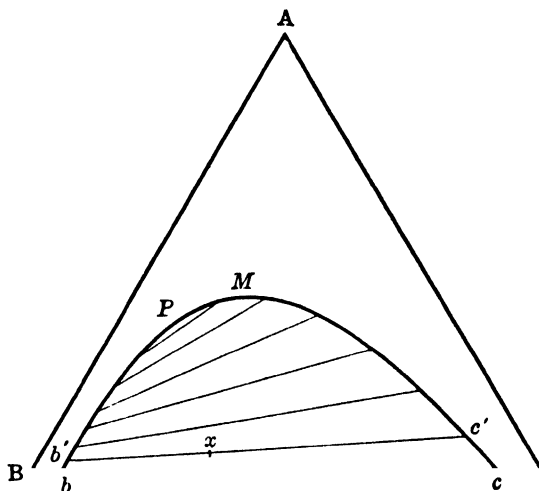


FIG. 204. One pair of partially miscible liquids

more of this component but, as is to be expected, the mutual solubilities of B and C are increased; * the compositions of the two layers therefore approach each other, as may be seen by the shortening of the tie-lines. At the point P the two conjugate solutions have the same composition, so that the two layers have become one; this is called the **critical point** of the system, or the **plait point**. The latter term is to be preferred since the expression "critical point" generally refers to a temperature, whereas in the case under consideration the temperature is fixed, and the so-called critical point refers to a particular composition.

When the various points representing the compositions of the conjugate layers are joined a **binodal curve**, bPc , is obtained, with a maximum at M . The latter does not in general coincide with the plait point P ;

* The distribution ratio of A between B and C does not remain constant, because of the increase in the mutual solubilities of B and C; the distribution law holds only if the nature of the two layers remains unchanged.

in fact it would only do so if the component A distributed itself equally between the two layers. Since, as a rule, A is more soluble in one or other of the layers, the tie-lines are not horizontal and P lies to one side of the maximum. Any point outside the curve $bPMc$ represents one liquid layer only, for by the phase rule the three-component system has then two degrees of freedom, apart from temperature and pressure, and the system can only be defined completely by fixing the compositions of two components. Inside the binodal curve $bPMc$, two liquid layers are in equilibrium, the compositions being given by the appropriate points on the curve; such systems have only one degree of freedom, at constant temperature and pressure, and consequently one concentration term, which allows the position of one layer to be fixed on the binodal curve, defines the system completely. In accordance with the properties of the triangular diagram, a mixture of the composition x , for example, will consist of the layers b' and c' , which are the extremities of the tie-lines passing through x ; the amounts of the layers are in the ratio of xc' to xb' , respectively.

Two general methods are available for obtaining the data for plotting the binodal curve for a ternary system together with some of the tie-lines. The most-obvious procedure is to make up various mixtures of the three components, allow them to come to equilibrium at a given temperature, and then to analyze the separate layers. This is not always convenient, and so another method is often used. Known amounts of two components are mixed and quantities of the third are added until the system previously consisting of two layers becomes homogeneous, or *vice versa*. The quantities of the three components for which this occurs give one point on the binodal curve; by means of a series of such points the whole curve can be plotted. To determine the positions of the tie-lines, definite amounts of the three liquids are taken, e.g., corresponding to the point x in Fig. 204. One of the layers is then separated, after equilibrium has been attained, and analysis for only one component allows the point b' to be fixed, since it must lie on the curve already plotted; if b' is then joined to x , the composition of the other layer c' is obtained by extrapolation, since this also must fall on the curve.

By means of binodal curves information can be obtained concerning the properties of liquid systems of the type being considered. For example, if initially a mixture of B and C, such as r , is taken (Fig. 205) and increasing amounts of A added, the composition of the whole system will follow the line rA , in accordance with the properties of the diagram. It is seen from the points at which rA cuts the tie-lines that the proportion of the left-hand layer, i.e., the one richer in B, becomes less and less, until at the point s it disappears completely and only one layer is left; between s and A the system consists of one liquid phase only. The same general phenomena will be obtained for any mixture of B and C except one: if the original composition is at t , so that the line tA passes through the plait point P , then instead of one layer disappearing steadily, the two layers remain but their compositions approach each other and become identical at P . The formation of one liquid phase at P is thus in a

sense different from the production of one layer at any other point on the curve.

The effects to be expected when B is added to mixtures of A and C, or when C is added to A and B, can be deduced in a similar manner and need not be described. Mention may, however, be made of the changes observed when the proportion of A is kept constant and the amounts of the other two components varied, e.g., along the line xy . At x the system consists of one liquid layer, but when the binodal curve is reached two layers commence to form; as the composition of the system moves towards y the relative amount of the left-hand layer decreases and that of the right-hand layer increases, as may be seen from the varying proportions in which the tie-lines are cut. When the right-hand side of the binodal curve is at-

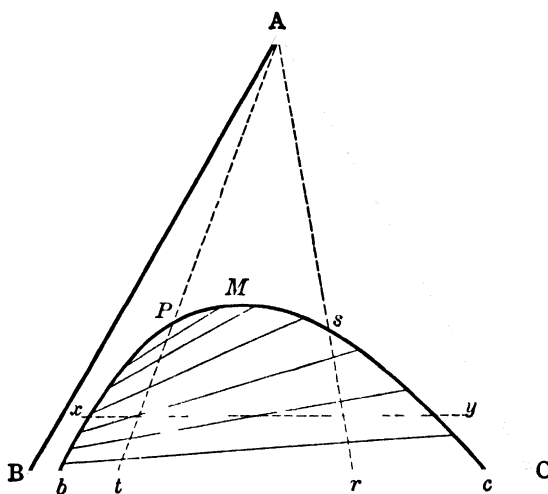


FIG. 205. Properties of system with one pair of partially miscible liquids

tained, the left-hand layer has completely disappeared, and once more the system has only one liquid phase. This change from one layer to two, and then back to one layer would be observed for any liquid containing less of the component A than corresponds to the maximum point M of the curve. If the original composition happened to lie between P and M then the conversion of one liquid layer to two would be seen to occur for a system which always contains more of the component A than at the critical point P ; such solutions are said to show **retrograde solubility**.

The Parkes Process.—A number of examples of three component systems in which only one pair of liquids is partially miscible have been studied, and the results have been applied in various theoretical and practical connections. An unusual system of this type, worthy of mention, is that formed by lead, silver and zinc in the liquid state; of these lead and silver, and zinc and silver are completely miscible, but lead and

zinc are only partially miscible. When the three metals are melted together, therefore, two liquid layers are formed, one consisting mainly of lead and the other mostly of zinc. It happens that silver is very much more soluble in the liquid rich in zinc, and so the ternary conjugate liquids are an upper layer, containing mainly zinc with a large proportion of silver and a little lead, and a lower, heavier layer consisting mostly of lead with a little silver and zinc. The Parkes process for the de-silverization of lead is based on this fact. Zinc is added to the molten mixture of lead and silver, and the two layers described above are formed. Nearly the whole of the silver then passes into the upper conjugate layer and as this has the higher melting point it solidifies first and so can be readily separated.

Influence of Temperature.—The binodal curves hitherto described are isothermal curves and their shape will alter with temperature, de-

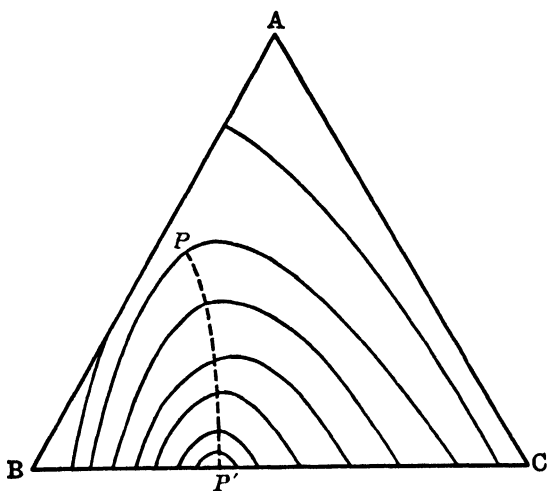


FIG. 206. Influence of temperature

pending on the changes in the mutual solubilities of the three components. Since cases of increased solubility with rising temperature are more common, these only will be considered. In order to represent the influence of temperature adequately a triangular prism should be used, the temperature being measured vertically, and a section at any particular temperature would then give a triangular diagram of the type already discussed. For present purposes, it is not necessary to draw the solid model, for the general behavior can be understood by plotting a number of isothermal binodal curves on one plane as in Fig. 206, which represents a common type of behavior. As the temperature increases the mutual solubilities of the components increase, and the curves enclose a smaller area. At a low temperature the area covered by a curve may become so

large that it cuts into the side AB ; this means that the two components A and B , as well as B and C , are now partially miscible, e.g., water, aniline and phenol at 50°C . The type of system into which that for one partially miscible pair may merge as the temperature is lowered will be considered shortly.

The plait points for the various temperatures fall on a curve PP' ; at temperatures above that of P' the two layers have ceased to exist, and the three liquid components are miscible in all proportions. It might appear, therefore, that P' represents the critical solution temperature for the ternary system; this is, however, not strictly true, for at P' the system consists of B and C only, and this is actually the *binary* critical solution temperature. When the binodal curves are of the form shown in Fig. 207, as is the case for the system water-acetone-phenol (F. A. H.

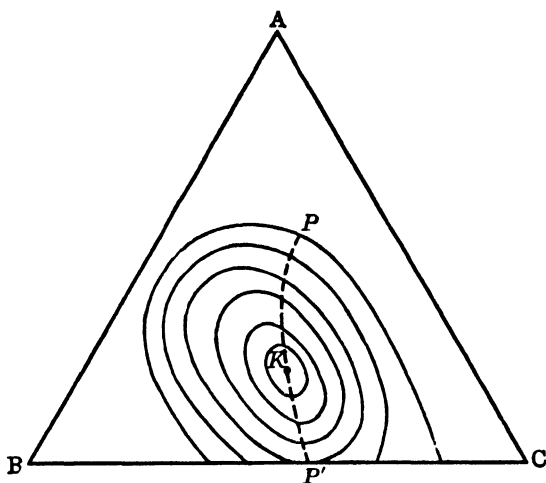


FIG. 207. Ternary system with critical solution temperature

Schreinemakers, 1900), then PKP' is the plait-point curve; K is the true ternary critical solution point. Above this temperature the three liquids are miscible in all proportions. The binary critical solution temperature for B and C is that at which the plait point, viz., P' , just touches the line BC ; it is seen that above this temperature B and C when alone are completely miscible, although addition of a certain amount of A may result in partial miscibility. It will be noted that the binodal curves between the temperatures P' and K are closed curves, and there are two plait points for each temperature.

(ii) *Two Pairs of Partially Miscible Liquids*.—When the liquids A and B , as well as B and C , are partially miscible, there may be two binodal curves, as in Fig. 208, each with its own tie-lines and plait point; this occurs with water, ethyl alcohol and succinic nitrile between 13°

and 31°C . It is unnecessary to discuss the properties of these curves as they are identical with those already considered. As the temperature is lowered the areas of partial miscibility generally increase and the two

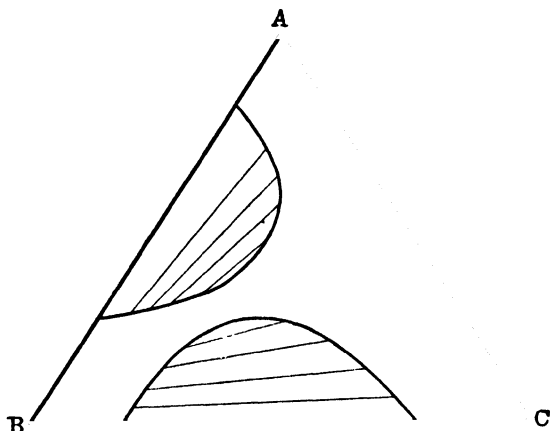


FIG. 208. Two pairs of partially miscible liquids

curves may coalesce to form a continuous band, as in Fig. 209; this is the same type of behavior as sometimes occurs with a system in which only one pair of liquids is normally partially miscible. If the binodal curve

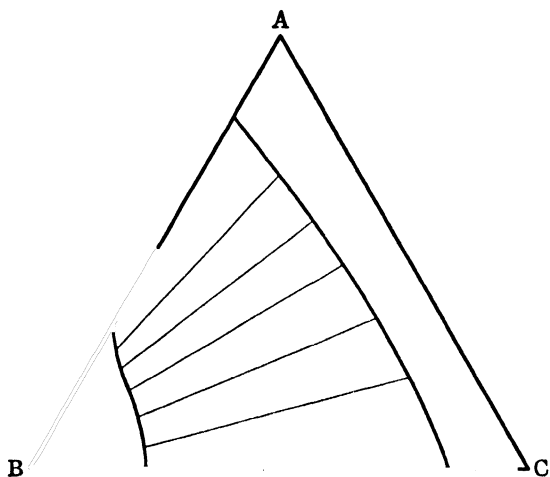


FIG. 209. Binoidal curve forming a band

forms a band, therefore, it is not possible to say whether on raising the temperature it will split up into one or two curves of the parabola-like shape. Several examples of three-component systems forming a binodal band are known, of which may be mentioned those involving

ethyl acetate, water and *n*-, *iso*-, or *sec*-butyl alcohol at 0° and 20° c. Any point inside the band represents two liquid layers, whereas outside there is only one liquid phase.

(iii) *Three Pairs of Partially Miscible Liquids*.—If none of the three

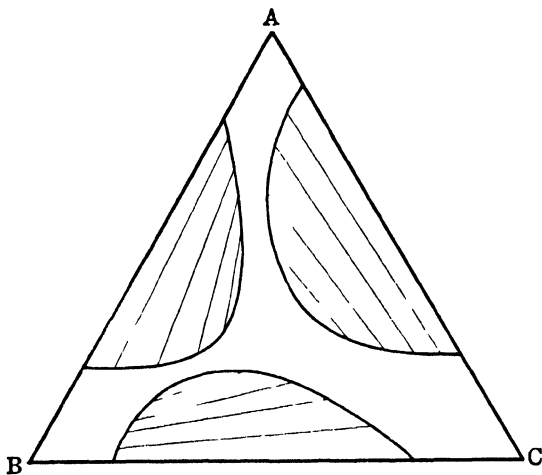


FIG. 210. Three pairs of partially miscible liquids

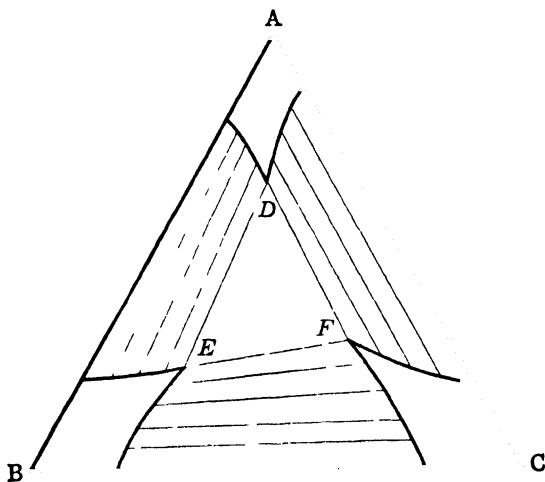


FIG. 211. Coalescence of three binodal curves

pairs of liquids is miscible, then three separate binodal curves are possible, points within each representing a pair of conjugate ternary liquids in equilibrium (Fig. 210). For less miscible liquids, or at lower temperatures, the curves may meet and the isothermal equilibrium diagram may be as shown in Fig. 211; the clear areas at the corners of the triangle repre-

sent one liquid phase only, whereas any point within the areas traversed by tie-lines implies two conjugate liquids whose compositions are given by the extremities of the tie-line passing through the particular point. The triangular area DEF is of special interest, for it is here that the three binodal curves coalesce, and any point in this area represents three liquid layers in equilibrium, their compositions being given by the points D , E and F . For three phases, ignoring vapor, the phase rule indicates two degrees of freedom for the three-component system; since temperature and pressure are fixed the system is invariant. Hence, when three liquid layers are in equilibrium their compositions must be definite, at a given temperature, irrespective of their relative amounts. The points D , E and F in Fig. 211 represent these definite concentrations. One of the best known systems giving a phase diagram of this type is water, ether and succinic nitrile (F. A. H. Schreinemakers, 1898).²⁷

Systems with Solid Phases.—In order to simplify the treatment it will be assumed that the components are only miscible in the liquid state, and the conditions under which solids separate from the liquid will be considered. It

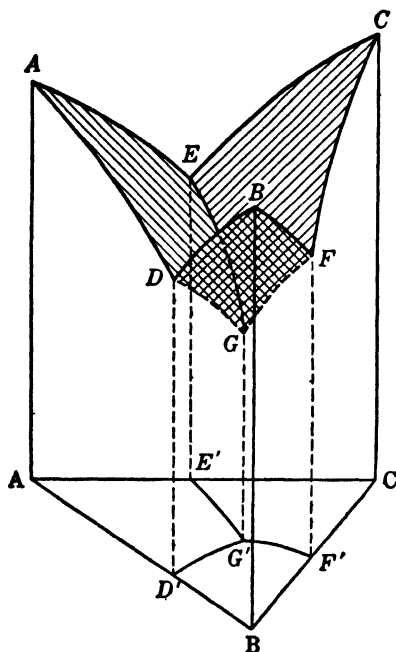


FIG. 212. Three component system with solid phases

will be supposed that the three pairs of solids, A and B, B and C, and C and A, do not combine nor do they form solid solutions, so that each pair gives a simple eutectic curve for the equilibrium between liquid and solid phases. It will be necessary, in order to understand the planar method of representation, to consider first the use of a solid model, as in Fig. 212; the compositions of various mixtures are given by the usual equilateral triangle, and temperatures are measured vertically. Each vertical face of the triangular prism represents a two-component system; thus for A and B, the solid-liquid equilibrium (freezing-point) curve is ADB , and for the binary systems B and C, and A and C, the equilibrium curves are BFC and CEA , respectively. The points D , E and F are the binary eutectics, at which two solid phases can exist in contact with binary liquid. If now the ternary system is considered, the phase rule shows that the conditions of equilibrium of liquid and one solid phase must be given

by a *surface* in the diagram; to describe the system, therefore, two variables still remain, viz., temperature and the proportion of one component. The three surfaces $AEGD$, $BFGD$ and $CFGE$ represent the conditions under which the solids A, B and C, respectively, separate from ternary liquid. The lines along

which two surfaces meet are the ternary eutectic curves, and they indicate the conditions for the separation of two solid phases from liquid; thus, along DG , EG , and FG the solids are A and B, B and C, and A and C, respectively. These three lines join at G , which is the ternary eutectic point, the only temperature at which the liquid phase can be in equilibrium with three solids, for application of the phase rule shows this to be an invariant point for the condensed system. Below the temperature G the ternary liquid cannot exist.

In order to represent the equilibrium conditions in two dimensions, the three binary and one ternary eutectic points and the curves joining them are first projected on to a horizontal plane, as shown at the base of the prism. Subsequently, curves are plotted on the same triangle representing sections through the solid model at a number of definite temperatures; in this way a diagram such as that in Fig. 213 is obtained. As before, D , E and F are the binary, and

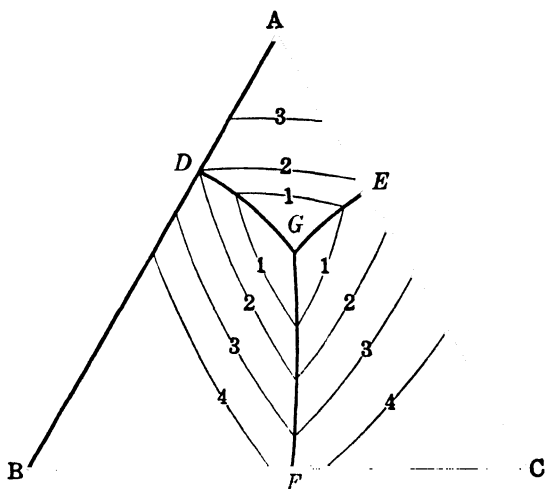


FIG. 213. Projection of phase diagram of three component system

G the ternary, eutectics; the curves DG , EG and FG give the conditions for the separation of two solid phases. The thin curves marked by numerals are isothermal sections through the surfaces on the solid model, and so represent one solid phase, viz., the component A, B or C occupying the corner nearest to the curve. As the temperature is increased the isothermal sections are indicated by 1, 2, 3, 4, etc.; the temperature of 2 is above the eutectic point E , but below the others; that for 3 is above the eutectics D and E , whereas 4 is for a temperature above these two eutectics and also above the melting point of A. It will be noted that at the temperature 4 the section of the curve in the A corner does not appear, since solid A does not then exist.

Systems of the type under consideration are of great importance in the study of metallic alloys and for the investigation of ceramic materials and glasses. The conditions are rendered complicated by the formation of polymorphic modifications, of compounds and solid solutions, and the whole subject is too complex to be considered further.

Two Salts and Water.—One of the most interesting types of three-component system involving solid phases, from the chemical point of view, arises when one of the components is water and the others are two salts *with an ion in common*. The melting point of water is very much lower than that of the other two components, and it is convenient to study such systems isothermally and to plot the curves representing the conditions of equilibrium of solid salt and ternary liquid. In other words, the solubility of each salt is determined in aqueous solutions, containing varying amounts of both, at constant temperature and atmospheric pressure.

The experimental method used is to make up various mixtures of the two salts which have a common ion, otherwise the system would be one of four components, and to dissolve them in water at a temperature slightly above that at which the system is to be studied. The solution is then placed in a thermostat at the required temperature and well stirred or shaken until equilibrium is attained; this may take a few hours only, but in some cases several days may be necessary, especially if there is a tendency for traces of gelatinous products to be formed by hydrolysis. Some of the clear liquid is then drawn off and analyzed. It is important also to know the nature of the solid phase; this is usually ascertained by removing some of the *wet* solid and analyzing it, and then determining by extrapolation the composition the dry solid would have. The procedure is generally known as the **residue method** of F. A. H. Schreinmakers (1893). If R in Fig. 214 is a point on the solubility

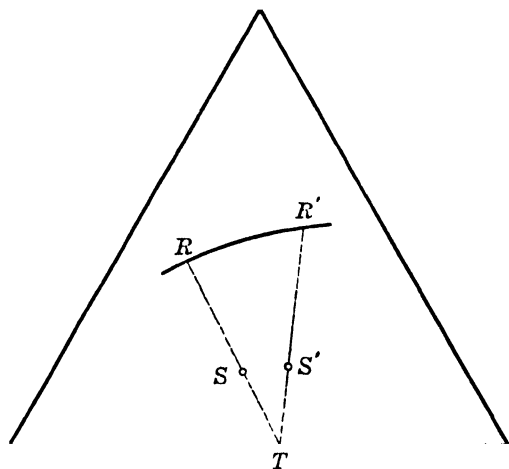


FIG. 214. The Schreinmakers residue method

curve, giving the composition of the liquid phase from which solid separates, and S is the composition of the *wet* solid, then the dry solid must be represented by a point on the line passing through R and S , since the wet solid may be regarded as a mixture of dry solid and solution (R) from which it has separated. The actual point can be found by considering another solution R' and the wet

solid S' separated from it, and continuing the line $R'S'$. The two lines meet at T , which gives the composition of the dry solid.

Several types of diagram have been used for depicting the conditions of equilibrium between solution and solid phases at a definite temperature in the systems made up of two salts and water. In the van't Hoff method rectangular coordinates are employed to represent the composition of the saturated solution; the number of grams or moles of each of the two salts associated with 100 grams or moles of water in the solution are plotted along the two axes. The method of B. Jänecke (1906) also involves a rectangular diagram; the proportions of the two dissolved salts are plotted as abscissae and the amounts of water in the saturated solution associated with 100 parts of total salt give the ordinates. Most workers at the present time, however, use the triangular diagram already described; this method will be adopted here, especially as it shows how systems of two salts and water are related to three component systems in general.

When the two salts S_1 and S_2 do not form a compound the isothermal solubility curves, giving the compositions of the solutions containing S_1 and S_2 in different proportions in equilibrium with either solid S_1 or S_2 , are similar to those shown in Fig. 215. There is a definite resemblance to

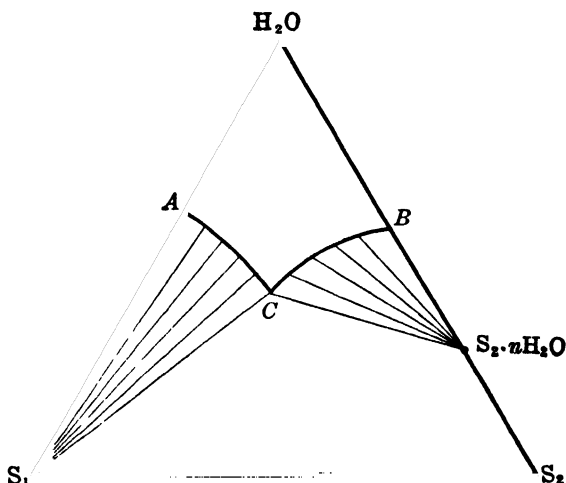


FIG. 215. Two salts and water

the curve for temperature 3 in Fig. 213, apart from the portion along which pure A separates. If the system of two salts and water were studied at temperatures below 0°C ., a corresponding curve representing the conditions for the separation of ice would have to be included; it is assumed in the present treatment, however, that the experimental temperature is always *above* the freezing point of water. The solubility curve in Fig. 215 is in two parts: along AC the solid phase is the salt S_1 , whereas along BC the solid salt S_2 is in equilibrium with solution. It should be evident from what has been said that the figure represents a horizontal section, at a definite temperature, through a space model like the one on

p. 800, so that the curves AC and BC are isothermal sections through the surfaces giving the conditions of equilibrium between ternary liquid and one solid phase. Where two surfaces in Fig. 212 meet there is a ternary eutectic line, and the point C in Fig. 215 is the section through this line; it represents the composition of liquid in equilibrium with the two solids, S_1 and S_2 , at the given temperature. It is not, however, called a eutectic, as it does not represent a temperature, but is referred to as an **isothermal invariant point**. At C there are, apart from vapor, three phases, viz., two solid and one liquid, and so the phase rule shows the three-component system to have two degrees of freedom; these are the pressure and temperature which have been fixed. The condensed system at C is, therefore, invariant, and the name is justified.

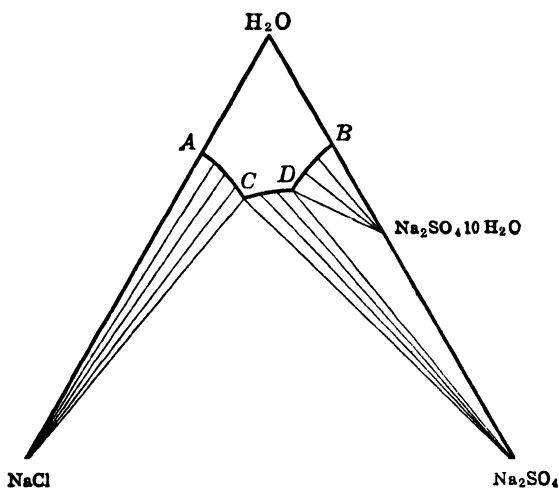


FIG. 216. The system $\text{NaCl-Na}_2\text{SO}_4\text{-H}_2\text{O}$ at 25°C .

In order to indicate the nature of the solid phase a number of points on the curves AC and BC , which give the compositions of the saturated solutions, are joined by tie-lines to points giving the compositions of the solid phases. In the figure, for example, it is seen that along AC the anhydrous salt S_1 separates, but the solutions given by points on BC are in equilibrium with a hydrate of S_2 . The tie-lines from AC meet in the S_1 corner, but those from BC meet at a point representing definite proportions of S_2 and H_2O , viz., $S_2 \cdot n\text{H}_2\text{O}$, where n is generally an integer. An example of this type of system is $\text{NaCl-Na}_2\text{SO}_4\text{-H}_2\text{O}$ at 15°C , the solid phases being NaCl and $\text{Na}_2\text{SO}_4 \cdot 10\text{H}_2\text{O}$. If both salts separate in the anhydrous form, then both sets of tie-lines meet in their respective corners, as is the case, for example, with $\text{NH}_4\text{Cl}-(\text{NH}_4)_2\text{SO}_4\text{-H}_2\text{O}$ at 30°C . In some cases both hydrated and anhydrous salts can be formed; this occurs with the $\text{NaCl-Na}_2\text{SO}_4\text{-H}_2\text{O}$ system between 17.9° and 32.4°C , the equilibrium diagram for 25°C being shown in Fig. 216. Along AC the solid

phase is NaCl , and along CD anhydrous Na_2SO_4 separates, whereas along DB the solution is in equilibrium with $\text{Na}_2\text{SO}_4 \cdot 10\text{H}_2\text{O}$. At the invariant points C and D , the two solids are NaCl and Na_2SO_4 , and Na_2SO_4 and $\text{Na}_2\text{SO}_4 \cdot 10\text{H}_2\text{O}$, respectively. It is of interest to observe that in the ternary system the anhydrous salt is stable at a temperature below the normal transition point for the binary system $\text{Na}_2\text{SO}_4\text{-H}_2\text{O}$; the ternary transition point, at which the three solids, Na_2SO_4 , $\text{Na}_2\text{SO}_4 \cdot 10\text{H}_2\text{O}$ and NaCl , can all be in equilibrium with solution is 17.9°C . Above 32.4°C ., the binary transition point, the decahydrate can no longer exist, and so at higher temperature there are only two parts to the solubility curve, the solid phases being NaCl and anhydrous Na_2SO_4 .

If, in addition to the pure components S_1 and S_2 , or their hydrates, a compound can separate from solution at the experimental temperature, the solubility curve will have an additional section CD , as in Fig. 217;

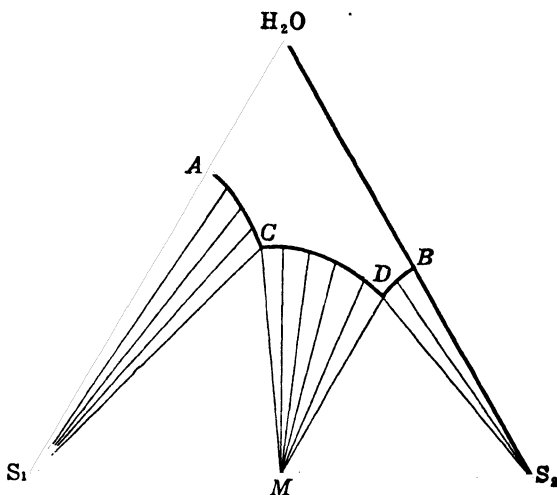


FIG. 217. Two salts and water: formation of a double salt

the composition of the compound is given by the point M . In this case it is seen that the compound, or double salt, like the single salts S_1 and S_2 , is anhydrous; if it were hydrated the tie-lines would have met in the interior of the triangle, as required for a phase containing all three components. The formation of two, or more, double salts will be indicated by the equivalent number of solubility curves, the tie-lines for which would meet in different points. The system $\text{NH}_4\text{NO}_3\text{-(NH}_4)_2\text{SO}_4\text{-H}_2\text{O}$ may be quoted in illustration, for two double salts, viz., $(\text{NH}_4)_2\text{SO}_4 \cdot 2\text{NH}_4\text{NO}_3$ and $(\text{NH}_4)_2\text{SO}_4 \cdot 3\text{NH}_4\text{NO}_3$, are obtained at 30°C . Many other cases of a similar type are known.

When the two salts form a continuous series of solid solutions, there is no isothermal invariant point, since not more than one solid phase can

be in equilibrium with liquid; the solubility curve, therefore, shows no break. Further, the composition of the solid phase varies with the composition of the solution from which it separates; this is shown by the spread of the tie-lines in Fig. 218. Behavior of this nature has been found in the system $\text{Ba}(\text{NO}_3)_2\text{-Pb}(\text{NO}_3)_2\text{-H}_2\text{O}$.

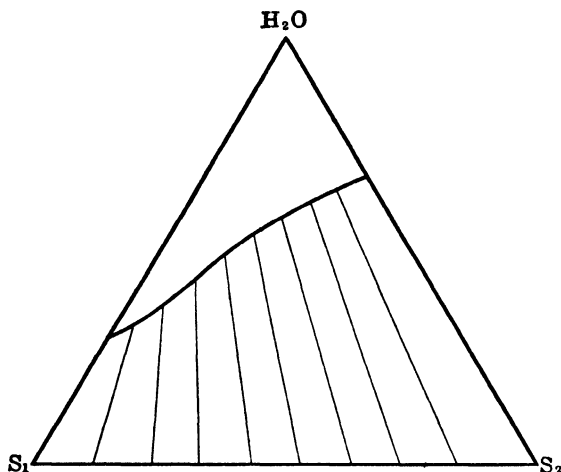


FIG. 218. Two salts and water: formation of solid solutions

Isothermal Evaporation.—Interesting conclusions concerning the behavior of various solutions can be reached from an examination of the equilibrium diagrams: the point *l* in Fig. 219, for example, represents an unsaturated solution, and if this is evaporated at constant temperature, the composition of the

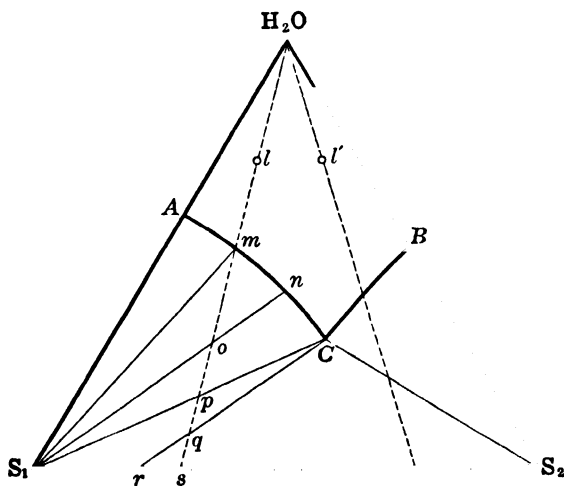


FIG. 219. Isothermal evaporation

system as a whole must travel along the dotted line drawn from the corner of the triangle and passing through l . When the solubility curve AC is reached at m , the solid S_1 begins to separate, with the result that the solution becomes richer in S_2 , and its composition must therefore follow the curve mC . When the composition of the whole system, for example, is at o , the solution is at n and the solid is S_1 ; this is in agreement with the fact that a system represented by a point on a tie-line consists of the two phases at the extremities of that line. At p the composition of the liquid phase will have reached C , the isothermal invariant point, when the solid S_2 commences to deposit in addition to S_1 . Since the condensed isothermal system has now no degrees of freedom, the composition of the liquid must remain at C while the two solids S_1 and S_2 separate out. For the system represented by q , for example, the solution is at C , and the solid will be a mixture of S_1 and S_2 of composition given by the point r . When s is finally reached the liquid phase will have disappeared, and the system will consist of the two solids only, in the proportion originally in the unsaturated solution. If the solution were initially at l' , then the behavior would be quite analogous, except for the fact that the solid S_2 would separate first. The other possibility to be considered is that the composition of the unsaturated solution lies on the line joining C to the apex of the triangle. On evaporation both solids would deposit simultaneously at C ; since the composition of the solution must remain unchanged, it follows that the proportions of S_1 and S_2 in the solid are the same as in the solution. The deposition of the mixture will continue until all the water has evaporated. A solution of this kind behaves as if it contained a single substance. The phenomena observed on the isothermal addition of water to any mixture of the salts will be the exact reverse of those seen on evaporation; starting from s , for example, the various stages will be passed through in the opposite order to those described above until l is reached.

When a double salt can exist as a solid phase the behavior on evaporation depends on whether the compound is stable in contact with water or not. In the example given in Fig. 220 the composition of the double salt is at M , and addition of water to it will cause the system to move up the line ME ; at the point E the solution is saturated with respect to double salt, since this is the solid phase in equilibrium with solution along curve CD . The addition of water, therefore, does not cause the double salt to decompose, but gives a stable saturated solution. In cases of this kind, it can be shown that from the standpoint of isothermal evaporation the whole system may be regarded as made up of two independent systems, viz., S_1 -M- H_2O and S_2 -M- H_2O , placed side by side, just as the equilibrium diagram of Fig. 179 (p. 756) may be considered as consisting of two separate portions. Solutions which on isothermal evaporation meet the solubility curves AC or EC will eventually form the invariant C , and this will finally dry up to give a mixture of solids S_1 and M . On the other hand, if the solubility curves ED and BD are met as water is removed, the solutions will change towards the invariant point D , and give on complete evaporation the solids S_2 and M . It is, therefore, unnecessary to discuss the case further, as it presents no new features.

If the solubility curve is of the form in Fig. 221, then the line joining the apex of the triangle to the point M giving the composition of the double salt cuts the solubility curve AC at m . This means that if water is added to M the resulting saturated solution is not in equilibrium with double salt, but with

the single salt S_1 ; in other words, addition of water brings about the decomposition of the double salt. If an unsaturated solution containing S_1 and S_2 in the proportions present in the double salt, e.g., as at l , is evaporated isothermally, the first solid to be deposited would not be the double salt, but the

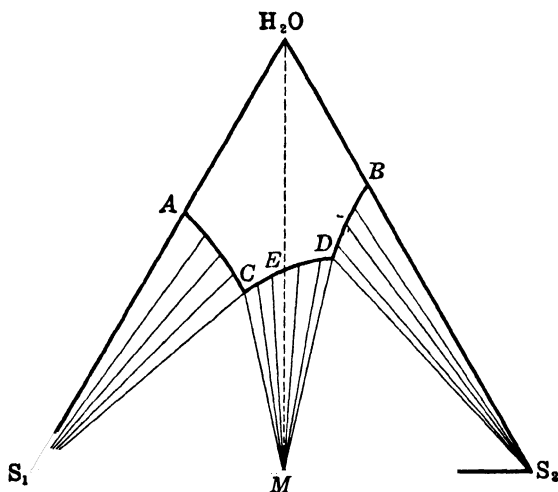


FIG. 220. Double salt stable in contact with water

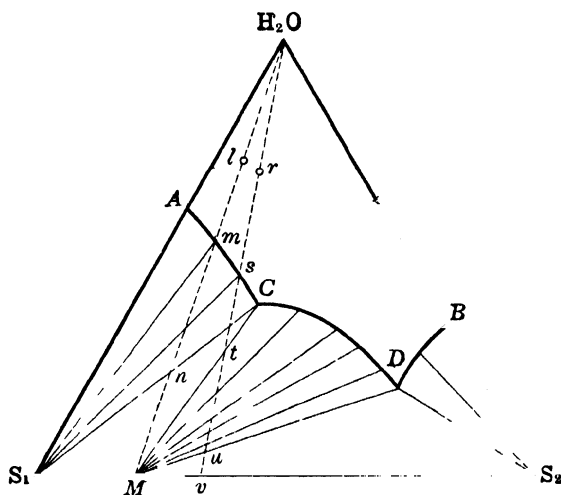


FIG. 221. Double salt unstable in contact with water

single salt S_1 . The whole system cannot now be regarded, as in the previous case, as made up of two independent systems, and the phenomena observed on evaporation are exceptional. Consider, for example, the solution l , in which the ratio of S_1 to S_2 is the same as in the double salt; as water is removed, solid

S_1 commences to be deposited at m , and the composition of the solution follows the solubility curve from m to C , as that of the whole system changes from m to n . When the solution has the composition of the invariant point C the double salt will separate; inspection of the diagram shows, however, that the ratio of S_1 to S_2 in the solution C is less than in the double salt, so that for the deposition of the latter to continue solid S_1 , previously separated, must redissolve. A solution behaving in this manner is said to be **incongruently saturated** (W. Meyerhoffer, 1897). As evaporation proceeds dissolution of S_1 and separation of double salt will continue, until eventually when all the water is removed only double salt will remain. The fact that the solid which has been deposited before the point C is reached must dissolve at C can be seen in another way. The composition of the solid phase in equilibrium with the invariant solution C is obtained by drawing tie-lines from C through the various points between n and M , representing the composition of the whole system (cf. Fig. 214). At n the solid phase is pure S_1 , whereas at the end of the evaporation it will be M , i.e., double salt; in between, therefore, the single salt S_1 must disappear, i.e., dissolve, so that the only double salt is left when the system is dry.

Unsaturated solutions lying to the left of l behave in a similar manner, the solution C being incongruently saturated; now, however, only part of the S_1 redissolves, leaving eventually a solid mixture of this and the double salt. If the unsaturated solution is at r , so that the solid S_1 first separates at s , between m and C , the solid S_1 will redissolve at C as usual, but when this phase has completely disappeared the composition of the whole system will be at t , and evaporation can still continue. Since there is now only one solid phase, viz., double salt, in contact with solution, the system is no longer invariant, and the composition of the liquid phase moves from C to D as that of the whole system changes from t to u , the double salt separating all the time. At D the system is again invariant and the composition of the solution remains constant while a mixture of double salt and S_2 is deposited; when evaporation is complete the solid phase has the composition given by v . The solution at D is thus congruently saturated, and evaporation proceeds quite normally of solutions of such concentration that they commence to deposit either double salt along CD or the single salt S_2 along BD .²⁸

Transition Point for Double Salt Formation.—Although two single salts do not yield a double salt under given conditions, they may do so as the temperature is raised or lowered; for example, in the system *d*-potassium tartrate, *d*-sodium tartrate and water, only single salts separate from solution above the transition point 55° c., but below this temperature the double salt, i.e., Rochelle salt, can be obtained. On the other hand, the racemic double salt between sodium ammonium *d*-tartrate and sodium ammonium *l*-tartrate can exist in contact with solution only above the transition point, 27° c.; it is of interest that Pasteur's success in separating the *d*- and *l*-forms of this salt was due to the chance that the crystals deposited below 27° c. Another example of double salt formation above the transition point occurs in the system $\text{MgSO}_4 \cdot 7\text{H}_2\text{O} \cdot \text{Na}_2\text{SO}_4 \cdot 10\text{H}_2\text{O} \cdot \text{H}_2\text{O}$, where the compound $\text{Na}_2\text{SO}_4 \cdot \text{MgSO}_4 \cdot 4\text{H}_2\text{O}$, i.e., *astracanite*, is in equilibrium with solution above 22° c. only. In general, if the double salt contains more water of crystallization than the two single salts the latter

are stable above, and the former below, the transition point, and *vice versa*; this is a direct result of the principle of mobile equilibrium (p. 831), since the heat evolved in the hydration of a salt is the main factor determining the heat of transition.

The relationship between a double salt and its constituent single salts at various temperatures may best be studied with the aid of isothermal solubility diagrams. In Fig. 222 AC and BC represent the normal solu-

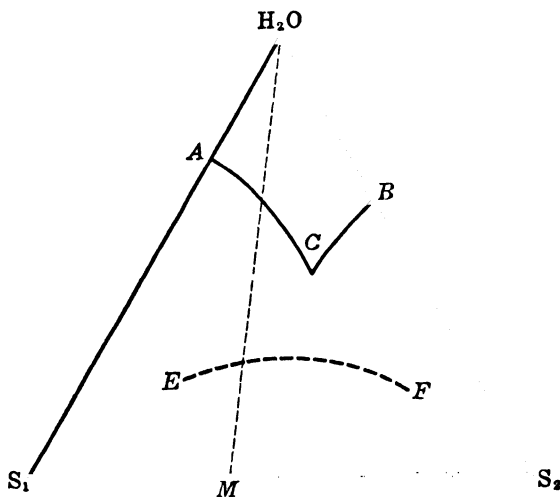


FIG. 222. Solubility curves outside transition interval

bility curves of the salts S_1 and S_2 , whereas EF is the imaginary, or metastable, curve for the double salt, giving the compositions of solutions in equilibrium with solid double salt. Since the curve EF would correspond to much more concentrated solutions than ACB , it obviously represents a metastable condition, and at the temperature for which the diagram applies the double salt is unstable. Suppose the temperature is now changed, either raised or lowered as required, so that the two curves approach. The condition in Fig. 223 will soon be reached; this represents the transition point, for at C the double salt first commences to be stable in contact with solution. The point C is here a true invariant, not merely an isothermal invariant, point for the condensed system; now three solid phases S_1 , S_2 and double salt, as well as solution, are in equilibrium. A further change of temperature leads to an isothermal diagram as in Fig. 224; the double salt, whose composition is represented by M , is now stable in contact with solution, but not in contact with water, as is shown by the line from the apex to M cutting the curve AC along which S_1 separates. Still further alteration of temperature leads to the curves in Fig. 225, where double salt is just stable in contact with water, and finally

curves such as those in Fig. 220, already considered, would be obtained. The range between Figs. 223 and 225, that is from the temperature at which double salt commences to separate from solution to that at which

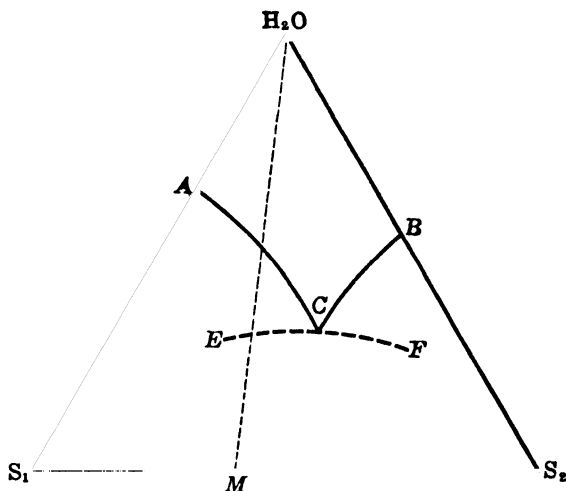


FIG. 223. Solubility curves at the transition point

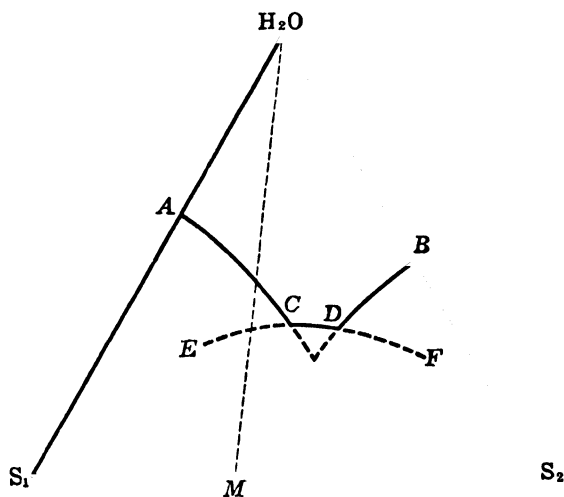


FIG. 224. Solubility curves in transition interval

it is stable in contact with water, is called the **transition interval**. An understanding of these facts has helped very materially in the elucidation of the properties of double salts and their solutions.

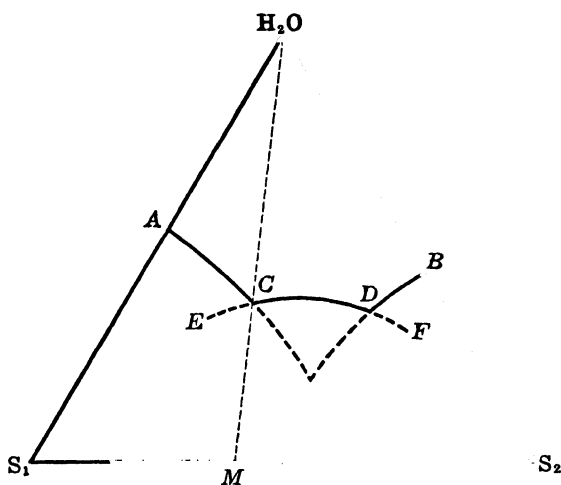
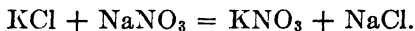


FIG. 225. Solubility curve at end of transition interval

Acid and Basic Salts.—Although the above treatment has referred particularly to double salts, the methods described are equally applicable for considering the conditions of stability and properties of acid and basic salts in aqueous solution. The three components may be taken as water, normal salt and its acidic, or basic, constituent, although the common practice is to choose the anhydride of the acid and the base, from which the salt is formed, as the two components in addition to water. For example, in the study of the formation of the series of basic bismuth nitrates, the three components are water, Bi_2O_3 and N_2O_5 ; this procedure would permit the representation of acid salts, if they existed, in addition to basic salts, on the same equilibrium diagram.

Four Component Systems.—Systems involving four components, particularly those with three salts having an ion in common, and water, are of industrial importance, but their treatment is outside the scope of this book. One special type of four-component system is, however, worthy of brief reference; it is the **reciprocal salt-pair** system (J. H. van't Hoff, 1895), which arises when two salts capable of undergoing double decomposition are dissolved in water, e.g.,



Although four simple salts are present in solution they constitute only three components, since the amount of the fourth is related to that of the other three by means of a chemical equation. In accordance with the definition of component on p. 474, it is evident that any three of the four substances involved in the double decomposition can be chosen as components of the system. The water in which the salts are dissolved constitutes the fourth component. For a discussion of such systems, the

reader must be referred to the special monographs dealing with phase equilibria.

REFERENCES

- Blasdale, W. C. "Equilibria in Saturated Salt Solutions."
 Bowden, S. T. "The Phase Rule and Phase Reactions."
 Clibbens, D. C. "The Principles of the Phase Theory."
 Findlay, A. and Campbell, A. N. "The Phase Rule and its Applications."
 Hildebrand, J. H. "Solubility."
 Rivett, A. C. D. "The Phase Rule."
 Roozeboom, H. W. B. "Phasenlehre."
 Taylor, H. S. "Treatise on Physical Chemistry," 2nd. ed., Chap. IX (A. E. Hill).
 Timmermans, J. "Chemical Species." (Translated by R. E. Oesper).
1. Sackur and Stern, *Z. Elek.*, **18**, 641 (1912); Sander, *Z. phys. Chem.*, **78**, 513 (1912); Lannung, *J. Am. Chem. Soc.*, **52**, 68 (1930); Wiebe, *et al.*, *ibid.*, **55**, 947 (1933); **56**, 76, 2357 (1934); Calingaert and Huggins, *ibid.*, **45**, 915 (1923); Bancroft and Belden, *J. Phys. Chem.*, **34**, 2123 (1930); Markham and Kobe, *Chem. Rev.*, **28**, 519 (1941).
 2. Philip, *J. Chem. Soc.*, **91**, 711 (1907); Debye and McAulay, *Phys. Z.*, **26**, 22 (1925); McAulay, *J. Phys. Chem.*, **30**, 1202 (1926); Butler, *ibid.*, **33**, 1015 (1929); Randall and Failey, *Chem. Rev.*, **4**, 285 (1927); Gross, *ibid.*, **13**, 91 (1933); Albright and Williams, *Trans. Farad. Soc.*, **33**, 247 (1937); Albright, *J. Am. Chem. Soc.*, **59**, 2098 (1937).
 3. Hildebrand, *ibid.*, **38**, 1452 (1916); *op. cit.*
 4. Ferguson and Funnell, *J. Phys. Chem.*, **33**, 1 (1929); Calingaert and Hitchcock, *J. Am. Chem. Soc.*, **49**, 750 (1927); Hovorka and Dreisbach, *ibid.*, **56**, 1664 (1934).
 5. Zawidski, *Z. phys. Chem.*, **35**, 129 (1900); Rosanoff, *J. Am. Chem. Soc.*, **36**, 1083 (1914); Sameshima, *ibid.*, **40**, 1489 (1918); Smyth and Engle, *ibid.*, **51**, 2646 (1929); Scatchard, *et al.*, *ibid.*, **60**, 1275 (1938); Lacher and Hunt, *ibid.*, **63**, 1752 (1941).
 6. Linebarger, *ibid.*, **17**, 615 (1895); Gahl, *Z. phys. Chem.*, **33**, 178 (1900); Bates and Kirschman, *J. Am. Chem. Soc.*, **41**, 1991 (1919); Thomas and Ramsay, *J. Chem. Soc.*, **123**, 3256 (1923); Dodson and Masson, *ibid.*, **125**, 668 (1924); Dunn and Rideal, *ibid.*, p. 676.
 7. Duhem, *Compt. rend.*, **102**, 1449 (1886); Margules, *Sitzungsber. Wien. Akad.*, **104**, 1243 (1895); Porter, *Trans. Farad. Soc.*, **16**, 336 (1920); Lewis and Murphee, *J. Am. Chem. Soc.*, **46**, 1 (1924).
 8. Smyth and Engle, ref. (5); Gilmann and Gross, *J. Am. Chem. Soc.*, **60**, 1525 (1938).
 9. Konowaloff, *Ann. Phys.*, **14**, 34 (1881).
 10. Tyrer, *J. Chem. Soc.*, **101**, 81 (1912); Othmer, *Ind. Eng. Chem.*, **20**, 743 (1928); Rosanoff, *et al.*, *J. Am. Chem. Soc.*, **31**, 448, 953 (1909); **37**, 301, 1072 (1915).
 11. Wade and Merriman, *J. Chem. Soc.*, **89**, 1656 (1905); **101**, 997 (1911); Merriman, *ibid.*, **103**, 628, 1790, 1801 (1913).
 12. Hulett and Bonner, *J. Am. Chem. Soc.*, **31**, 390 (1909); Foulk and Hollingsworth, *ibid.*, **45**, 1223 (1923); Bonner, *et al.*, *ibid.*, **52**, 633, 1747 (1930); Ewell and Welch, *ibid.*, **63**, 2475 (1941).
 13. Alexejeff, *Bull. Soc. chim.*, **38**, 145 (1882); *Ann. Phys. Chim.*, **28**, 305 (1886); Hill, *J. Am. Chem. Soc.*, **45**, 1143, 1924 (1923); **46**, 918 (1924); **48**, 918 (1926); Keyes and Hildebrand, *ibid.*, **39**, 2126 (1917); Rothmund, *Z. phys. Chem.*, **26**, 443 (1898); Ludeman, *Ind. Eng. Chem. (Anal.)*, **12**, 446 (1940); Francis, *Ind. Eng. Chem.*, **33**, 554 (1941).
 14. Hudson, *Z. phys. Chem.*, **47**, 113 (1904); Cox, *J. Am. Chem. Soc.*, **48**, 451 (1926); **49**, 1080 (1927).
 15. Timmermans, *Z. phys. Chem.*, **58**, 129 (1907); Jones, *J. Chem. Soc.*, **123**, 1374 (1923); Glasstone, *et al.*, *ibid.*, 635 (1927); Patterson, *ibid.*, 1745 (1937); Lucasse, *et al.*, *J. Phys. Chem.*, **46**, 500 (1942); **47**, 183 (1943); see also, ref. (2).

16. Marshall, *J. Chem. Soc.*, **89**, 1350 (1906); Merriman, *ibid.*, **103**, 1790 (1913); Mains, *Chem. Met. Eng.*, **26**, 779 (1922); van Klooster and Douglas *J. Phys. Chem.*, **49**, 67 (1945).
17. Berthelot and Jungfleisch, *Ann. Chim. Phys.*, **26**, 396 (1872); Nernst, *Z. phys. Chem.*, **8**, 110 (1891); Jakowkin, *ibid.*, **13**, 359 (1894); **18**, 585 (1895); **20**, 19 (1896); **29**, 613 (1899); Kuriloff, *ibid.*, **24**, 697 (1897); **25**, 419 (1898); Dawson, *J. Chem. Soc.*, **77**, 1239 (1900); **79**, 238, 1072 (1901); **89**, 1666 (1906); Moore, *et al.*, *ibid.*, **101**, 1635 (1912); 1447 (1931); Wynne-Jones, *ibid.*, 1064 (1930); Linhart, *J. Am. Chem. Soc.*, **37**, 258 (1915); **40**, 158 (1918); Hill, *ibid.*, **43**, 254 (1921); **46**, 1132 (1924).
18. Hildebrand, *ref.* (3); *J. Am. Chem. Soc.*, **39**, 2297 (1917); **42**, 2180 (1920); *op. cit.*
19. Guthrie, *Phil. Mag.*, **17**, 462 (1884).
20. Rheinboldt, *et al.*, *J. prak. Chem.*, **111**, 242 (1925); Pounder and Masson, *J. Chem. Soc.*, 1357 (1934).
21. Bancroft, *J. Phys. Chem.*, **2**, 143, 245 (1898); **3**, 144 (1899).
22. Kremann, *Monatsh.*, **25**, 1216 (1904); van Laar, *Z. phys. Chem.*, **63**, 216 (1908); **64**, 257; **66**, 197 (1909); Kendall and Booge, *J. Chem. Soc.*, 1768 (1925); Ross and Sommerville, *ibid.*, 2770 (1926).
23. Ropzeboom, *Z. phys. Chem.*, **10**, 477 (1892).
24. Wilson, *J. Am. Chem. Soc.*, **43**, 719 (1921); Carpenter and Jette, *ibid.*, **45**, 578 (1923); Partington and Dolan, *Trans. Farad. Soc.*, **20**, 342 (1924); Collins and Menzies, *J. Phys. Chem.*, **40**, 379 (1936).
25. Gillespie, *et al.*, *J. Am. Chem. Soc.*, **48**, 1209 (1926); **58**, 2565 (1936); **61**, 2496 (1939); Ubbelohde, *Trans. Farad. Soc.*, **28**, 275, 282 (1932); Owen and Jones, *Proc. Phys. Soc.*, **49**, 587, 603 (1937).
26. Keevil, *J. Am. Chem. Soc.*, **64**, 841 (1942).
27. Wright, *Proc. Roy. Soc.*, **49A**, 174 (1891); **50A**, 372 (1892); Schreinmakers, *Z. phys. Chem.*, **27**, 205 (1898); **33**, 78 (1900).
28. Meyerhoffer, *Ber.*, **30**, 1810 (1897).

CHAPTER XI

CHEMICAL EQUILIBRIUM AND FREE ENERGY

THE EQUILIBRIUM CONSTANT

Introduction.—The earliest attempts to account for the tendency for a particular reaction to occur were expressed in terms of the **affinity** of the reacting substance. The word *affinitas* appears to have been first used in this connection by Albertus Magnus (c. 1250), and the subject of chemical affinity was discussed by a number of well-known scientists of the 17th century, e.g., R. Boyle, I. Newton and J. R. Glauber. In 1718 the first "tables of affinity," in which substances were arranged in a series giving the order of replacement from chemical combination, was compiled by E. F. Geoffroy, and this method of representing affinity was extended by T. Bergman (1775) and others. As far back as 1674, Boyle had realized that the *quantity* of reactant material was of importance and that large amounts could "compensate for its want of strength," and Bergman had noted that an excess of reacting substance was frequently necessary to make a given process go to completion. An attempt was made by C. F. Wenzel (1777) to measure the affinity between substances by the rate at which they reacted, and as a result of an investigation of the action of acids on metals he found the rate of solution to be proportional to the concentration of acid.

It is seen, therefore, that there was a growing realization of the influence of the amounts of the reacting substances on the velocity and extent of a chemical reaction, but it is to C. L. Berthollet (1799) that the credit must be given for the first clear emphasis on the importance of mass or concentration. He showed that whereas in the laboratory sodium carbonate and calcium chloride reacted to form calcium carbonate and sodium chloride, the reverse process was occurring in certain Egyptian lakes on the shores of which sodium carbonate was actually crystallizing out. This reversal of the usual reaction was attributed to the large amount of sodium chloride present in the lake water. Berthollet also showed that the extent of decomposition of barium sulfate and calcium carbonate by potassium hydroxide solution depended on the quantity of the latter reagent used, and that by employing a large excess the insoluble salts could be almost completely decomposed. The conclusion based on these experiments was stated by Berthollet (1803) in the following way: "the chemical activity of a substance depends on the force of its affinity and upon the mass which is present in a given volume."

It was unfortunate that this important hypothesis was rejected by most workers at the beginning of the 19th century, because its author was

of the opinion that the amount of reactant not only affected the extent of the reaction but also the composition of the product. The establishment of the law of constant proportions at about the same time showed the latter part of Berthollet's contention to be incorrect, and so the former was disregarded, although the eminent chemists J. L. Gay-Lussac and J. J. Berzelius appear to have accepted the view that mass was important in chemical reactions. The work of H. Rose (1842), on the decomposition of alkaline-earth sulfides by water and on the reversible reaction between a soluble carbonate and an insoluble sulfate, once more centered attention on the influence of mass, and in 1850 L. Wilhelmy, in the course of a study of the inversion of sucrose in the presence of acid, showed that the rate of the reaction at any instant was proportional to the amount of sucrose remaining unchanged at that instant. The effect of mass on chemical reaction was also brought out clearly by M. Berthelot and P. St. Gilles (1862-63) who studied the reversible formation of ester from ethyl alcohol and acetic acid; they found that the reaction did not go to completion, but the proportion of acid converted into ester increased as the amount of alcohol was increased. For example, when equimolar amounts of alcohol and acid were employed only 66 per cent of the latter appeared as ester, but if the proportion of alcohol was increased four-fold about 88 per cent of the acid was converted into ester. Berthelot and St. Gilles also observed that the rate of formation of ester in any mixture of alcohol and acid was proportional to the products of the masses of the reactants, and inversely proportional to the total volume; in other words, the rate of reaction was found to be approximately proportional to the concentrations of the reactants, but this important result was not generalized.

Although it was commonly realized prior to this that many reactions were reversible, in the sense that they could occur in either direction according to the experimental conditions, and that a state of equilibrium was attained in which the reaction was not complete, it was considered that the equilibrium was static and that in this state all reaction ceased. As a consequence of his interpretation of the mechanism of the etherification of alcohol by concentrated sulfuric acid, A. W. Williamson (1850) was led to the view that chemical equilibrium was *dynamic*, and the reaction appeared to be stationary because both direct and reverse processes were taking place at the same rate. This conception of equilibrium was supported by F. G. M. Malaguti (1857) and by L. Pfaundler (1857) and it is, of course, now universally accepted. It appeared to C. M. Guldberg and P. Waage (1864-67) that a reversible reaction involving dynamic equilibrium provided the best conditions for the comparison of chemical affinities, and this led them to state what is now known as the **law of mass action** and to formulate mathematically the influence of mass on chemical activity.

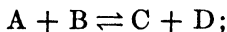
Law of Mass Action.—The fundamental postulate of Guldberg and Waage was that, for a homogeneous system, *the rate of a chemical reaction is proportional to the active masses of the reacting substances*, the molecular

concentration of a substance in solution or in the gas phase being taken as a measure of its active mass. From this they developed a method for the comparison of affinities by assuming that the proportionality constant was the "affinity coefficient" for the given reactants; thus if two substances A and B take part in a reversible reaction in which the products are A' and B', and the active masses at equilibrium are p and q , and p' and q' , respectively, then the rates of the direct and reverse reactions are kpq and $k'p'q'$, respectively, where k and k' are the "affinity coefficients." Since the two rates are equal at equilibrium it follows that

$$kpq = k'p'q', \quad (1)$$

and hence Guldberg and Waage concluded that "by determining experimentally the active masses p , q , p' and q' (at equilibrium) there can be found the relation between the affinity coefficients k and k' ." In a later paper Guldberg and Waage (1879) discarded the term affinity coefficient and used "coefficient of velocity" for k and k' ; this corresponds to the modern interpretation of these constants (p. 1046). The proportionality constants were thus no longer regarded as indicating the affinities of the reactions, but the other aspects of the law of mass action remained unaltered. Guldberg and Waage pointed out, for example, that if the ratio k/k' is known for any reaction "we can calculate the result of the reaction for any original condition of the four substances," and it is this aspect of their law which will be considered in some detail.

The Equilibrium Constant.—Consider the reversible homogeneous gas reaction



then the rate of the direct reaction, i.e., left to right, is kc_{ACB} , assuming the concentrations, represented by the c terms, to give a measure of the active masses, and the rate of the reverse reaction is $k'c_{CCD}$. At equilibrium the rates of the two reactions will be equal; it follows, therefore, that if the concentrations are those at equilibrium

$$kc_{ACB} = k'c_{CCD}; \quad (2)$$

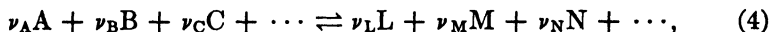
$$\therefore \frac{c_{CCD}}{c_{ACB}} = \frac{k}{k'} = K, \quad (3)$$

the constant K being called the **equilibrium constant** of the reaction.* This equation, derived from the law of mass action, gives the condition of equilibrium of a reversible chemical reaction; it has been referred to as the expression of the **law of chemical equilibrium**. No matter what are the initial concentrations of A, B, C and D, when equilibrium is attained the concentrations must be related to one another in such a way that (3) holds. It is evident, as Guldberg and Waage pointed out, that

* The reciprocal of (3) is, of course, also a constant, but the convention universally adopted in expressing the equilibrium constant is to put the concentrations of the products, i.e., the right-hand side of the chemical equation, in the numerator.

provided K is known it is possible to calculate the position of equilibrium for any initial concentrations of the four reacting substances; this aspect of the subject will be discussed later.

The reaction considered above involved a single molecule of each of two reactants and two products; in practice, however, other types of reactions occur, and the following general case may be examined:



where ν_A , ν_B , etc., represent the number of molecules of reactants A, B, etc., and ν_L , ν_M , etc., the number of molecules of products L, M, etc. According to Guldberg and Waage the rate of such a reaction is proportional to the active mass, i.e., concentration, of each species raised to the power of the number of molecules taking part in the reaction as expressed by the equation; it follows, therefore, that at equilibrium

$$k(c_A)^{\nu_A}(c_B)^{\nu_B}\cdots = k'(c_L)^{\nu_L}(c_M)^{\nu_M}\cdots; \quad (5)$$

$$\therefore \frac{(c_L)^{\nu_L}(c_M)^{\nu_M}}{(c_A)^{\nu_A}(c_B)^{\nu_B}\cdots} = \frac{k}{k'} = K. \quad (6)$$

Thermodynamic Derivation of the Law of Chemical Equilibrium.—

The law of chemical equilibrium may also be obtained by thermodynamic methods, as was shown by A. Horstmann (1873), J. Willard Gibbs (1876) and J. H. van't Hoff (1886). The derivation given by van't Hoff is as follows, for the gas reaction $A + B \rightleftharpoons C + D$. Imagine two large vessels, at the same temperature, in each of which the four species A, B, C, D are always in equilibrium; suppose, further, that each of four walls of the vessels is permeable to only one of the substances.* Let the partial pressures in the first vessel be p_A , p_B , p_C and p_D and in the second vessel p'_A , p'_B , p'_C and p'_D , the species A, B, C and D being indicated by subscripts. The assumption is made that the vessels are so large that the transfer of relatively small quantities of material from one to the other results in no appreciable change in the partial pressures or concentrations of the substances taking part in the equilibrium. Further, *it is assumed that all the reacting substances are ideal gases.*

An isothermal and reversible process is now carried out. By means of the walls permeable to A, one mole of A from the first vessel is transferred isothermally and reversibly to the second; the initial pressure is p_A and the final pressure p'_A , so that the increase of free energy of the substance A is [cf. equation (129), p. 233]

$$\Delta F_A = \int_{p_A}^{p'_A} \nu_A dp. \quad (7)$$

Since the reacting substances are ideal gases, this becomes

$$\Delta F_A = RT \ln p'_A/p_A \quad (8)$$

* A hypothetical vessel of this kind is often referred to as van't Hoff's "equilibrium box"; the concept is due to van't Hoff, but the name was apparently suggested by F. Haber.

Similarly, if one mole of B is transferred in the same direction the free energy change is given by

$$\Delta F_B = RT \ln p'_B/p_B. \quad (9)$$

Imagine that at the same time as A and B are transferred in one direction, one mole of C and one of D are transferred isothermally and reversibly in the opposite direction; the accompanying free energy changes are

$$\Delta F_C = RT \ln p_C/p'_C \quad \text{and} \quad \Delta F_D = RT \ln p_D/p'_D. \quad (10)$$

The total change of free energy is the sum of these four terms, and so

$$\Delta F = RT \ln \frac{p'_A}{p_A} + RT \ln \frac{p'_B}{p_B} + RT \ln \frac{p_C}{p'_C} + RT \ln \frac{p_D}{p'_D}. \quad (11)$$

Since the vessels are so large that the systems remain at equilibrium, the net change of free energy at constant temperature will be zero, i.e., $\Delta F = 0$, and hence, from (11),

$$\ln \frac{p'_A}{p_A} + \ln \frac{p'_B}{p_B} = \ln \frac{p'_C}{p_C} + \ln \frac{p'_D}{p_D}; \quad (12)$$

$$\therefore \frac{p_C \times p_D}{p_A \times p_B} = \frac{p'_C \times p'_D}{p'_A \times p'_B} = \text{constant } (K_p). \quad (13)$$

Since the partial pressures in each vessel refer to equilibrium conditions, it follows that the product of the partial pressures of C and D, divided by the product for A and B, is a constant at equilibrium at definite temperature; this constant is represented by K_p , and applies for a system *involving ideal gases*. For such gases the partial pressures are proportional to the molar concentrations (p. 303); therefore (13) becomes

$$\frac{c_C \times c_D}{c_A \times c_B} = \text{constant } (K_c), \quad (14)$$

where the c terms are the equilibrium concentrations. This result is identical with (3), and so the law of chemical equilibrium may be regarded as established thermodynamically for a system in which all the reacting species are ideal gases. It will be evident, too, that the equilibrium equation can then be expressed either in terms of concentrations or of partial pressures.

If instead of taking a simple reaction, a more general one, such as that represented by (4), had been considered, the corresponding equilibrium constant, viz., equation (6), would have resulted; it is only necessary to remember that the free energy change for the transfer of ν moles is ν times the value for 1 mole. Instead of giving the deduction in detail, however, an alternative proof of the law of equilibrium, based on the use of chemical potentials, will be described.

Consider the system of equation (4) in a state of equilibrium; suppose a small variation is made in the system, so that the numbers of moles of the reactants A, B, etc., are decreased by $\nu_A dx$, $\nu_B dx$, etc., whereas the numbers of moles of products L, M, etc., are correspondingly increased by $\nu_L dx$, $\nu_M dx$, etc. The temperature and pressure are assumed to remain constant while the small change takes place, and so the corresponding free energy change (cf. p. 238) is

$$(\partial F)_{T,P} = \mu_L(\nu_L dx) + \mu_M(\nu_M dx) + \cdots - \mu_A(\nu_A dx) - \mu_B(\nu_B dx) - \cdots, \quad (15)$$

where μ_A , μ_B , etc., are the chemical potentials of A, B, etc., and μ_L , μ_M , etc., are those of L, M, etc. Since the system is at equilibrium a small change at constant temperature and pressure results in no change in the free energy (p. 234). The right-hand side of (15) may, therefore, be put equal to zero; hence,

$$\nu_L \mu_L + \nu_M \mu_M + \cdots = \nu_A \mu_A + \nu_B \mu_B + \cdots. \quad (16)$$

This is the general thermodynamic condition for equilibrium in a reversible chemical reaction; it is applicable to systems of all types. It may be written in the alternative form

$$(\nu_L \mu_L + \nu_M \mu_M + \cdots) - (\nu_A \mu_A + \nu_B \mu_B + \cdots) = \Sigma \nu \mu = 0. \quad (17)$$

The chemical potential of any component of an *ideal gas* mixture may be written (p. 304) as

$$\mu = \mu_p^0 + RT \ln p,$$

where p is the partial pressure of the given gas in the mixture, and μ_p^0 is a constant for that gas depending only on temperature. Substituting the appropriate values in (16) it follows that

$$\begin{aligned} \nu_L RT \ln p_L + \nu_M RT \ln p_M + \cdots + \Sigma \nu' \mu_p^0 \\ = \nu_A RT \ln p_A + \nu_B RT \ln p_B + \cdots + \Sigma \nu \mu_p^0, \end{aligned} \quad (18)$$

where the $\Sigma \nu \mu_p^0$ terms are constant; hence,

$$\frac{(p_L)^{\nu_L} (p_M)^{\nu_M} \cdots}{(p_A)^{\nu_A} (p_B)^{\nu_B} \cdots} = \text{constant } (K_p) \quad (19)$$

for an ideal gaseous system at constant temperature. Alternatively, the chemical potential for an ideal gas may be written

$$\mu = \mu_c^0 + RT \ln c,$$

the constant μ_c^0 being a function of temperature, and by an exactly analogous method to that given above it follows that

$$\frac{(c_L)^{\nu_L} (c_M)^{\nu_M} \cdots}{(c_A)^{\nu_A} (c_B)^{\nu_B} \cdots} = \text{constant } (K_c) \quad (20)$$

at a definite temperature, so that the law of equilibrium may be regarded as proved in the general case. Further, it is possible to express the

chemical potential of a component in an ideal gas mixture in the form

$$\mu = \mu_x^0 + RT \ln x,$$

where x is the mole fraction and μ_x^0 depends on both temperature and pressure (p. 305); if the corresponding values are introduced in (16), it is seen that

$$\frac{(x_L)^{\nu_L}(x_M)^{\nu_M}\cdots}{(x_A)^{\nu_A}(x_B)^{\nu_B}\cdots} = \text{constant } (K_x) \quad (21)$$

at constant temperature and pressure. It is apparent that there are three different ways of expressing the equilibrium constant for a system of ideal gases, viz., in terms of either partial pressures, concentrations or mole fractions. It is of interest to compare the values of the three constants; this may be done by writing $p = RTc$ and substituting in (19). The result is

$$K_p = \frac{(c_L)^{\nu_L}(c_M)^{\nu_M}\cdots(RT)^{\nu_L+\nu_M+\cdots}}{(c_A)^{\nu_A}(c_B)^{\nu_B}\cdots(RT)^{\nu_A+\nu_B+\cdots}} = K_c(RT)^{[(\nu_L+\nu_M+\cdots)-(\nu_A+\nu_B+\cdots)]} \quad (22)$$

or

$$K_p = K_c(RT)^{\Delta\nu}, \quad (23)$$

where $\Delta\nu = (\nu_L + \nu_M + \cdots) - (\nu_A + \nu_B + \cdots)$, is equal to the increase in the number of molecules in the reaction as represented by the stoichiometric equation (4). Only when $\Delta\nu$ is zero, that is, for a process in which the number of molecules of reactants and products are equal, are K_p and K_c , involving partial pressures and molar concentrations respectively, identical.

The partial pressure p of any gas in a mixture of ideal gases is related to the total pressure P by the expression $p = xP$ (see p. 303), and substitution in (19) gives

$$K_p = K_x P^{\Delta\nu} \quad (24)$$

and from (23),

$$K_c = K_x (P/RT)^{\Delta\nu} = K_x V^{\Delta\nu}, \quad (25)$$

where V is the volume of the system containing 1 mole of gas, i.e., the total volume divided by the total number of moles present. If $\Delta\nu$ is zero, all three equilibrium constants are identical, but otherwise they must differ. It will be observed that (24) and (25) relating K_x to the other constants involve the total pressure or volume of the system. It follows, therefore, as indicated above, that whereas K_p and K_c vary only with temperature, the value of K_x , when $\Delta\nu$ is not zero, depends also on the total pressure; for this reason K_x is not often used in connection with gas reactions.

Equilibrium in Liquid Systems.—In the deduction of (16) no assumptions is involved as to the nature of the constituents, and it applies equally to any homogeneous system, liquid as well as gaseous. The chemical potential of any constituent of an ideal liquid mixture, for which Raoult's

law or Henry's law holds, is given by (cf. p. 684)

$$\mu = \mu_x^0 + RT \ln x, \quad (26)$$

where x is the mole fraction of any constituent in the mixture, and μ_x^0 depends on temperature and pressure. It follows, therefore, by substitution in (16) that

$$K_z = \frac{(x_L)^{r_L}(x_M)^{r_M} \cdots}{(x_A)^{r_A}(x_B)^{r_B} \cdots}. \quad (27)$$

This equation for the equilibrium constant holds at constant temperature and pressure both for a system in which all the liquids take part in the reaction, e.g., ethyl alcohol, acetic acid, ethyl acetate and water, and also for one in which the reacting substances are dissolved in an inert solvent; in the latter case the solvent must, of course, be taken into account when evaluating the mole fractions of the substances involved in the reaction. If n_A , n_B , etc., and n_L , n_M , etc., are the numbers of moles of reactants and products, having molecular weights M_A , M_B , etc., and M_L , M_M , etc., respectively, and n_0 is the number of moles of solvent of molecular weight M_0 , then

$$x_A = \frac{n_A}{n_0 + n_A + n_B + \cdots + n_L + n_M + \cdots} = \frac{n_A}{\Sigma n}. \quad (28)$$

The concentration, in moles per liter, of the same component is obtained by dividing n_A by the volume of the system in liters, and the latter can be expressed in terms of the total mass of the constituents of the solution and its density ρ ; thus,

$$\begin{aligned} c_A &= \frac{1000n_A\rho}{(n_0M_0 + n_AM_A + n_BM_B + \cdots + n_LM_L + n_MM_M + \cdots)} \\ &= \frac{1000n_A\rho}{\Sigma nM}; \end{aligned} \quad (29)$$

$$\therefore x_A = \frac{c_A}{1000\rho} \cdot \frac{\Sigma nM}{\Sigma n}. \quad (30)$$

If the solution is dilute the number of moles of solvent is greatly in excess of the total number of moles of the reacting species, so that

$$\begin{aligned} \Sigma nM &\approx n_0M_0 \quad \text{and} \quad \Sigma n \approx n_0; \\ \therefore x_A &\approx \frac{M_0}{1000\rho} c_A. \end{aligned} \quad (31)$$

For a dilute solution ρ is practically equal to the density of the solvent, and since this and its molecular weight M_0 are constant, it follows that the concentration of any constituent in a dilute solution is approximately proportional to its mole fraction. It is, therefore, possible to express the equilibrium constant for a reaction in a homogeneous liquid phase in the

form

$$K_c = \frac{(c_L)^{\nu_L}(c_M)^{\nu_M} \dots}{(c_A)^{\nu_A}(c_B)^{\nu_B} \dots}, \quad (32)$$

the concentrations being in moles per unit volume, e.g., per liter, provided the reacting species are all present in dilute solution. Theoretically both K_x and K_c for a reaction in a liquid system should depend on the pressure, but the actual variation is small, and can generally be neglected.

Influence of Inert Gases and Excess of Reactants.—For the general homogeneous gas reaction (4) the equilibrium constant K_p may be written

$$K_p = \frac{(n_L)^{\nu_L}(n_M)^{\nu_M} \dots}{(n_A)^{\nu_A}(n_B)^{\nu_B} \dots} \left(\frac{P}{\Sigma n} \right)^{\Delta \nu} \quad (33)$$

by combining (21), (24) and (27). If an inert gas is added Σn is increased, but if $\Delta \nu$ is zero the position of equilibrium will remain unaffected. If $\Delta \nu$ is positive, that is, there are more molecules on the right-hand side than on the left-hand side of the chemical equation, and the pressure is kept constant when the inert gas is added, i.e., the volume is increased, then $(\Sigma n)^{\Delta \nu}$ will increase, and in order that K_p may remain constant the equilibrium will be displaced in such a way that n_A , n_B , etc., will be decreased and n_L , n_M , etc., increased. That is to say, the inert gas will cause the reaction to go from left to right. On the other hand, if $\Delta \nu$ is negative the addition of an inert gas will shift the position of equilibrium from right to left, provided the pressure is maintained constant. If the volume is constant, an increase of Σn will result in an exactly proportional increase of P and the position of equilibrium will remain unchanged.

It is generally assumed that the addition of one of the substances on the left-hand side of the chemical equation will move the equilibrium to the right, and *vice versa*; this is only true if the volume of the system is kept unchanged. If the pressure remains constant, and the volume increases, it is possible that the shift of equilibrium will actually be in the reverse direction; this depends on the value and sign of $\Delta \nu$ and the number ν of molecules of the added substance involved in the reaction. Under constant pressure conditions, therefore, each case should be considered separately.

The Reaction Isotherm.—When considering a small change occurring in a reversible reaction (p. 820) it was assumed the system was at equilibrium, so that the free energy change could be equated to zero. If the constituents are, however, not at equilibrium but at certain arbitrary concentrations, there is a definite change of free energy as a result of the conversion of the reactants A, B, etc., to the products L, M, etc.; this may be determined in a simple manner by the hypothetical "equilibrium box." As before, imagine a large box in which the various gases taking part in the reaction are present at equilibrium, and let the partial pressures be $(p_A)_e$, $(p_B)_e$, etc., and $(p_L)_e$, $(p_M)_e$, etc., the subscript *e* indicating equilibrium conditions. Suppose ν_A moles of the reactant A at an

arbitrary pressure p_A are transferred isothermally and reversibly into the box through the wall permeable to A; the free energy change, assuming only ideal gases to be involved, is then

$$\Delta F_A = \nu_A RT \ln (p_A)_e / p_A. \quad (34)$$

Similarly, if ν_B moles of B are transferred at the same time the corresponding free energy change ΔF_B is given by an analogous expression, and so on for all the other reactants. Simultaneously with the passage of the reactants into the box it may be assumed that the products are withdrawn isothermally and reversibly, and when ν_L moles of L, ν_M moles of M, etc., have been removed the equilibrium box is restored exactly to its original state. The net increase of free energy for the whole process is

$$\Delta F = \left\{ \nu_A RT \ln \frac{(p_A)_e}{p_A} + \nu_B RT \ln \frac{(p_B)_e}{p_B} + \dots \right\} - \left\{ \nu_L RT \ln \frac{(p_L)_e}{p_L} + \nu_M RT \ln \frac{(p_M)_e}{p_M} + \dots \right\}. \quad (35)$$

This equation may be written in the form

$$-\Delta F = RT \ln \frac{(p_L)_e^{\nu_L} (p_M)_e^{\nu_M} \dots}{(p_A)_e^{\nu_A} (p_B)_e^{\nu_B} \dots} - RT \ln \frac{(p_L)^{\nu_L} (p_M)^{\nu_M} \dots}{(p_A)^{\nu_A} (p_B)^{\nu_B} \dots}. \quad (36)$$

The fraction involving the equilibrium partial pressures in the first term on the right-hand side is obviously equal to K_p , and the fraction in the second term is of the same form but it involves *arbitrary*, instead of equilibrium, partial pressures. The second term is generally written $RT \Sigma \nu \ln p$, where $\Sigma \nu \ln p$ represents the *algebraic* sum of all the $\nu \ln p$ terms, those for products being taken as positive and for reactants as negative. The free energy change for the transfer of reactants at arbitrary partial pressures to products at arbitrary pressures is then given by the expression

$$-\Delta F = RT \ln K_p - RT \Sigma \nu \ln p, \quad (37)$$

which is one form of the **reaction isotherm** first deduced by J. H. van't Hoff (1886). If all the p_e and p terms in (36) were replaced by the appropriate values of RTc , where c is the concentration of each gaseous reactant, assumed to be ideal, the additional RT terms cancel and there results

$$-\Delta F = RT \ln K_c - RT \Sigma \nu \ln c. \quad (38)$$

Similarly, a third form of the isotherm involving mole fractions can be readily derived, viz.,

$$-\Delta F = RT \ln K_x - RT \Sigma \nu \ln x \quad (39)$$

applicable at constant pressure. If the arbitrary partial pressures, or concentrations, of reactants and products happen to be those required for equilibrium, then $\ln K_p$ is equal to $\Sigma \nu \ln p$, and the free energy change

is zero. In other words, a reversible transfer from reactants to products at equilibrium results in no change in the free energy, at constant temperature. This is in agreement with the criterion of equilibrium previously derived (p. 234); the assumption is made, as on p. 818, that the extent of the system is so great that there is no appreciable change in the partial pressures or concentrations of the substances concerned.

Ideal Liquid Systems.—The change of free energy resulting from the transfer of reactants into products, all at arbitrary concentrations, in an ideal liquid system may be calculated in a similar manner to that described for gases. When one of the reacting species of an ideal liquid mixture, in which its mole fraction is x , is transferred isothermally and reversibly to a system at equilibrium in which its mole fraction is x_e , the increase of free energy per mole is equal to the difference of chemical potential in the two states; that is,

$$\Delta F = \mu_e - \mu = RT \ln x_e/x. \quad (40)$$

By the use of the equilibrium box device, the reactants at arbitrary concentrations being passed reversibly into the box while chemically equivalent amounts of products at arbitrary concentrations are withdrawn reversibly, it can be deduced that

$$-\Delta F = RT \ln K_x - RT \sum \nu \ln x. \quad (41)^*$$

This equation, like (27), is applicable when all the constituents of the liquid mixture take part in the reaction, and also when they are dissolved in an inert solvent, provided the system behaves ideally. If the solution is dilute then the form

$$-\Delta F = RT \ln K_c - RT \sum \nu \ln c \quad (42)$$

may be employed.

Nonideal Systems.—The foregoing deductions are all based on the assumption that the systems, whether gaseous or liquid, are ideal, and it is important that this should be remembered when applying the equations for the equilibrium constant in the forms given above. The same methods may be used, however, to derive equations of general applicability. Utilizing (16) or (17), as the condition of equilibrium, together with the expression (cf. p. 684)

$$\mu = \mu^0 + RT \ln a$$

for the chemical potential, it can be readily shown that for any system, whether ideal or not,

$$\frac{(a_L)^{\nu_L}(a_M)^{\nu_M} \dots}{(a_A)^{\nu_A}(a_B)^{\nu_B} \dots} = K, \quad (43)$$

the a terms representing the activities of the various reacting substances indicated by the subscripts. If the system is ideal, the activities may be replaced by the respective partial pressures, concentrations or mole

* It may be mentioned that (40) and (41) will hold if either Henry's law or Raoult's law is obeyed.

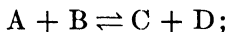
fractions of gaseous reactants, or the mole fractions of liquids. In dilute solution, concentrations may be used, since they are proportional to the latter. If the actual partial pressures, mole fractions or concentrations are employed in the expression for the law of chemical equilibrium, then the "equilibrium constant" for a nonideal system will vary with the total pressure or concentration, instead of being a true constant. The result so obtained is often called the "classical equilibrium constant." If the values are extrapolated to low pressure or to infinite dilution, the true, or thermodynamic, equilibrium constant, identical with that of (43), is obtained.

The general expression for the free energy analogous to (43) may be derived by the methods already given; it may be written as

$$-\Delta F = RT \ln K - RT \sum \nu \ln a. \quad (44)$$

This form should be used for all systems, particularly those that are not ideal, but for ideal systems it reduces to (37), (38) or (39) for gases, or to (41) for liquids. If the solution is dilute, as well as ideal, it becomes equivalent to (42).

Activities and Reaction Rates.—Since the correct expression for the true, or thermodynamic, equilibrium constant involves activities, and not concentrations or mole fractions, it appears at first sight that the velocity of a chemical reaction should depend on the activities, rather than on the concentrations, of the reacting substances. Consider, for example, the reversible reaction



then if the velocities of direct and reverse processes are proportional to the activities of the various species, it follows that at equilibrium $ka_Aa_B = k'a_Ca_D$, so that

$$\frac{a_Ca_D}{a_Aa_B} = \frac{k}{k'} = K, \quad (43a)$$

in agreement with (43). Since reaction depends on encounters between molecules, and the number of encounters depends on the *concentrations* of the reacting species, it is most probable that the rate of reaction is actually proportional to the molar concentrations of the reactants, as assumed by Guldberg and Waage. It is, however, necessary to introduce a **kinetic factor** \mathfrak{F} , depending on the nature of the medium in which the reaction occurs; thus,

$$\begin{aligned} \text{Velocity of direct reaction} &= k_1c_Ac_B\mathfrak{F}, \\ \text{Velocity of reverse reaction} &= k'_1c_Cc_D\mathfrak{F}', \end{aligned}$$

so that at equilibrium, $k_1c_Ac_B\mathfrak{F} = k'_1c_Cc_D\mathfrak{F}'$, and therefore

$$\frac{c_Cc_D}{c_Ac_B} \cdot \frac{\mathfrak{F}'}{\mathfrak{F}} = \frac{k_1}{k'_1} = K.$$

If the quantity $\mathfrak{F}'/\mathfrak{F}$ is equal to $f_C f_D / f_A f_B$, where the f terms are the activity coefficients of the appropriate species which vary with the medium, this equation becomes identical with (43a). It is thus possible to correlate the dependence of reaction velocity on concentrations with the use of activities in the equation for the equilibrium constant. If the reacting systems are ideal in behavior, then \mathfrak{F} and \mathfrak{F}' are unity, since activities and concentrations may then be taken as identical. Further consideration of the kinetic factors will be found in Chapter XIII.

Standard Free Energy Changes.—If all the substances taking part in a reaction are in their respective standard states of unit activity, the general form (44) for the reaction isotherm reduces to

$$\Delta F^0 = -RT \ln K, \quad (45)$$

for the **standard free energy change**, ΔF^0 , for the given reaction at the temperature T . It follows, therefore, that (44) may conveniently be written as

$$\Delta F = \Delta F^0 + RT \sum \nu \ln a. \quad (46)$$

For a system involving gases the standard state of each substance is the ideal gas at 1 atm. pressure, and the corresponding equilibrium constant is K_p , i.e.,

$$\Delta F_p^0 = -RT \ln K_p, \quad (47)$$

with the activities expressed in terms of pressures in atm. units. If the system behaves ideally actual pressures in atm. may be used, so that (46) becomes

$$\Delta F = \Delta F_p^0 + RT \sum \nu \ln p, \quad (48)$$

which is equivalent to (37). For a liquid mixture, the standard state of each constituent is the pure liquid, the activities being expressed in terms of mole fractions; then

$$\Delta F_x^0 = -RT \ln K_x, \quad (49)$$

and for an ideal system,

$$\Delta F = \Delta F_x^0 + RT \sum \nu \ln x, \quad (50)$$

which is the analogue of (41). For reactions taking place in dilute solution the standard state is chosen so that the activity is equal to the concentration or molality at infinite dilution (p. 685); the standard free energy change, in the former case, is given by

$$\Delta F_c^0 = -RT \ln K_c, \quad (51)$$

while the corresponding form of the reaction isotherm, assuming ideal behavior is,

$$\Delta F = \Delta F_c^0 + RT \sum \nu \ln c, \quad (52)$$

which is similar to (42).

The importance of the foregoing equations lies in the fact that the standard free energies of reactions, derived from the equilibrium con-

stants by (47), (49) or (51), according to the conditions, can be conveniently expressed analytically, as will be seen below. From these data it is then possible, by means of (48), (50) or (52), to calculate the free energy change for the reaction with pressures, mole fractions or concentrations, respectively, at any arbitrary values. Further reference to this matter will be made later (p. 853).

The van't Hoff Equation.—If (47) is differentiated with respect to temperature, at constant pressure, the result is

$$\left(\frac{\partial(\Delta F^0)}{\partial T} \right)_P = -R \ln K_p - RT \frac{d \ln K_p}{dT},$$

the partial differential notation of the final term being omitted since K_p is independent of pressure. Upon multiplying through by T and substituting ΔF^0 for $-RT \ln K_p$ on the right-hand side, it is seen that

$$T \left(\frac{\partial(\Delta F^0)}{\partial T} \right)_P = \Delta F^0 - RT^2 \frac{d \ln K_p}{dT}.$$

The Gibbs-Helmholtz equation (125), p. 232, for the special case in which the substances are all in their standard states, is

$$\Delta F^0 = \Delta H^0 + T \left(\frac{\partial(\Delta F^0)}{\partial T} \right)_P,$$

and combination with the preceding equation gives

$$\frac{d \ln K_p}{dT} = \frac{\Delta H^0}{RT^2}. \quad (53)$$

For an ideal gaseous system, ΔH^0 , the standard change in the heat content, may be replaced by ΔH , without specifying standard states. For a nonideal system the difference between the two quantities depends on the variation of the activities (or activity coefficients) with temperature (cf. p. 687); since the total effect for a reaction is usually small, it is possible to write in most cases

$$\frac{d \ln K_p}{dT} = \frac{\Delta H}{RT^2}. \quad (54)$$

This result, which is of fundamental importance in chemistry, represents the variation with temperature of the equilibrium constant for a reaction involving gases in terms of the change in heat content, that is, in terms of the heat of reaction at constant pressure. It has been referred to as the **reaction isochore** (Greek: *equal space*), because it was first deduced by van't Hoff for a constant volume system. Since ΔH is the heat of reaction at constant pressure, the name "isochore" is misleading, and hence (54) will be called the **van't Hoff equation**.

The alternative form of (54), involving the heat of reaction at constant volume, is obtained by taking logarithms of (23), that is,

$$\ln K_p = \ln K_c + \Delta\nu \ln RT. \quad (55)$$

Since R is constant, differentiation with respect to temperature gives

$$\frac{d \ln K_p}{dT} = \frac{d \ln K_c}{dT} + \frac{\Delta\nu}{T},$$

and upon introducing (54), the result is

$$\frac{d \ln K_c}{dT} = \frac{\Delta H}{RT^2} - \frac{\Delta\nu}{T} = \frac{\Delta H - RT\Delta\nu}{RT^2}. \quad (56)$$

Recalling that $\Delta\nu$ is the increase in the number of molecules in the gas reaction, it follows, assuming ideal behavior, that (cf. p. 201)

$$\Delta H - RT\Delta\nu = \Delta E,$$

and hence

$$\frac{d \ln K_c}{dT} = \frac{\Delta E}{RT^2}, \quad (57)$$

where the increase of energy ΔE is equal to the heat of reaction at constant volume.

A form of the van't Hoff equation involving K_x may be derived from (24) and (54). Since K_x varies with pressure it is necessary to indicate constant pressure conditions; thus,

$$\left(\frac{\partial \ln K_x}{\partial T} \right)_P = \frac{\Delta H}{RT^2}. \quad (58)$$

In this form the equation is also applicable to *ideal solutions*, as may be shown from (49) by employing the same method as for gases. A similar equation involving K_c , instead of K_x , may also be used for very dilute solutions which behave ideally, if the coefficient of thermal expansion is neglected.

Integration of the van't Hoff Equation.—In order to integrate any form of the van't Hoff equation it is necessary to express ΔH or ΔE as a function of temperature; this can be done (cf. p. 214) provided the variation with temperature of the heat capacities of reactants and products is known. Integration of the Kirchhoff equation (p. 213) leads to the result

$$\Delta H = \Delta H_0 + \int_0^T \Delta C_P dT, \quad (59)$$

where ΔC_P is the increase heat of capacity in the reaction as represented by the chemical equation. Substitution in (54) gives

$$\frac{d \ln K_p}{dT} = \frac{\Delta H_0}{RT^2} + \frac{1}{RT^2} \int_0^T \Delta C_P dT$$

and on integration between T_1 and T_2 ,

$$\ln \frac{(K_p)_2}{(K_p)_1} = -\frac{\Delta H_0}{R} \left(\frac{1}{T_2} - \frac{1}{T_1} \right) + \frac{1}{R} \int_{T_1}^{T_2} \frac{dT}{T^2} \int_0^T \Delta C_P dT. \quad (60)$$

The variation of ΔC_P with temperature may be represented by (cf. p. 214)

$$\Delta C_P = \alpha + \beta T + \gamma T^2 + \dots,$$

and upon substituting in (60), and carrying out the integrations, there results

$$\begin{aligned} \ln \frac{(K_p)_2}{(K_p)_1} = & -\frac{\Delta H_0}{R} \left(\frac{1}{T_2} - \frac{1}{T_1} \right) + \frac{\alpha}{R} \ln \frac{T_2}{T_1} + \frac{\beta}{2R} (T_2 - T_1) \\ & + \frac{\gamma}{6R} (T_2^3 - T_1^3) + \dots. \end{aligned} \quad (61)^*$$

In exactly the same way, general integration of (54) gives

$$\ln K_p = -\frac{\Delta H_0}{RT} + \frac{\alpha}{R} \ln T + \frac{\beta}{2R} T + \frac{\gamma}{6R} T^2 + \dots + \frac{I}{R}, \quad (62)$$

where I is an integration constant. For approximate purposes, or for temperatures which are close together, the value of ΔH may be assumed to be constant, and on integration of (54) the forms

$$\ln K_p = -\frac{\Delta H}{RT} + I', \quad (63)$$

I' being an integration constant, and

$$\ln \frac{(K_p)_2}{(K_p)_1} = -\frac{\Delta H}{R} \left(\frac{1}{T_2} - \frac{1}{T_1} \right)$$

or, with ΔH in calories, so that R is 1.987,

$$\log \frac{(K_p)_2}{(K_p)_1} = -\frac{\Delta H}{4.576} \left(\frac{1}{T_2} - \frac{1}{T_1} \right) \quad (64)$$

are obtained, ΔH being an average value for the temperature range considered. If ΔH is known from the heat of reaction, at constant pressure, then $(K_p)_2$ at any temperature T_2 can be calculated if the value $(K_p)_1$ at another temperature T_1 is known; alternatively if K_p is known at two temperatures ΔH can be determined. This method has been frequently used for the evaluation of heats of reaction; if accurate results are required it is essential to employ (61), but for most purposes, provided a small temperature range is involved, (63) and (64) give reasonably satisfactory values.

Since $-\Delta F^0$ for a gas reaction is equal to $RT \ln K_p$, by (47), it is evident that multiplication of (62) by RT gives an expression for ΔF^0 as a

* The same equation may be obtained by substituting (64), p. 214, directly into (54) above, and integrating.

function of temperature; thus,

$$-\Delta F^0 = -\Delta H^0 + \alpha T \ln T + \frac{\beta}{2} T^2 + \frac{\gamma}{6} T^3 + \cdots + IT. \quad (65)$$

If K_p (or ΔF^0) is known at two temperatures, the constants ΔH^0 and I for the given reaction can be evaluated, and since α , β , γ , etc., can be derived from heat capacity data, it is possible to calculate the standard free energy change at any temperature in the range over which the heat capacity data are applicable.

Principle of Mobile Equilibrium.—By means of the equilibrium constant and van't Hoff equations it is possible, as will be seen shortly, to determine quantitatively the influence of pressure and temperature on the position of equilibrium in a reversible reaction. The same conclusions may be reached in a qualitative manner by the **principle of mobile equilibrium**, developed independently by H. Le Chatelier (1885) and F. Braun (1886), which may be stated in the following form. *If a change occurs in one of the factors, such as temperature or pressure, under which a system is in equilibrium, the system will tend to adjust itself so as to annul, as far as possible, the effect of that change.* For example, if the pressure of a gas mixture at equilibrium is increased, that is, to say, the volume occupied by the molecules is diminished, the reaction will tend to occur involving a decrease in the number of molecules, and the equilibrium will be shifted in that direction; similarly, a diminution of pressure will result in a movement in the direction of an increased number of molecules. If the temperature is raised that reaction will occur in which heat is absorbed, whereas a lowering of temperature will move the equilibrium in the direction of the exothermic reaction. Actual examples of the application of the Le Chatelier-Braun principle are given below. It may be noted that the principle, which can be derived from the second law of thermodynamics, is of general applicability to all systems in equilibrium, e.g., solid-liquid, solid-solid, etc. (cf. p. 461), and is not necessarily restricted to chemical reactions.¹

Experimental Methods.—The experimental method adopted for the study of equilibria in gaseous systems depends chiefly on the nature of the reaction and on the temperature at which it is to be studied. If this is not too high the reacting substances may be sealed into glass bulbs and heated at the requisite temperature. After allowing time for equilibrium to be established, the bulbs are cooled rapidly so as to fix the equilibrium, and the contents are then analyzed. This procedure was adopted by M. Bodenstein (1894-99) in his classical work on the hydrogen-iodine reaction. He started with either pure hydrogen iodide or with mixtures of hydrogen and iodine, so that equilibrium was approached from both sides. After reaction had occurred the bulbs were broken under dilute alkali, if hydrogen iodide or an equimolar mixture of hydrogen and iodine were the initial materials, and the volume of the residual hydrogen gas was measured. Knowing the initial amounts of the reactants the partial pressures at equilibrium could be calculated. When the quantities of

hydrogen and iodine were not equivalent the hydrogen iodide and iodine in the equilibrium mixture were absorbed in water and both determined by analysis; the residual gas consisted of the unchanged hydrogen at equilibrium.

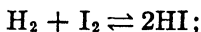
For reactions involving hydrogen a method has been used in which the partial pressure of this gas is measured with the aid of a semipermeable membrane. At high temperatures, platinum and iridium, as well as palladium (p. 664), are permeable to hydrogen but not to other gases. A bulb made of one of these metals, attached to a sensitive manometer, is inserted in the reacting gases and the increase of pressure observed is equal to the partial pressure of hydrogen in the system. The principle has been applied to study the dissociation of hydrogen sulfide, iodide and bromide, and also to determine the equilibrium constant in the reaction between hydrogen and sulfur dioxide.

If the reaction is accompanied by a change in the number of molecules, the position of equilibrium can often be determined by measuring the pressure of the system at constant volume; the special form of the Victor Meyer apparatus designed for high temperature work (p. 316) has been employed for this purpose, although it gives only approximate results. More accurate data have been obtained by the use of special glass and quartz-thread manometers sealed into the reaction vessel (p. 318). The effusion method for determining gas densities has also been adapted to the study of homogeneous equilibria, particularly those involving dissociation, e.g., bromine and iodine molecules into atoms.

A flow method, based on the "hot and cold tube" of H. St. Clair Deville (1864-66), has been employed, mainly by W. Nernst and his collaborators (1904-06). The reacting gases are passed through a tube heated to the desired temperature, in which equilibrium is attained. The gases are then cooled rapidly by passage through a capillary tube; they are subsequently collected and analyzed. The presence of a catalyst in the reaction tube facilitates the attainment of equilibrium without affecting its position (p. 1127), but it must be excluded from the colder parts of the apparatus, otherwise the equilibrium will be altered while the gases are cooling.

HOMOGENEOUS EQUILIBRIA

The Hydrogen-Iodine Reaction.—Gaseous reversible systems are best considered under two headings: those in which the number of molecules is unchanged, and those in which different numbers of molecules appear on the two sides of the chemical equation. A familiar example in the former category is the reaction



as seen on p. 821, for such processes K_p , K_c and K_x are identical, so that it is immaterial how the concentrations at equilibrium are expressed. Reactions of this type have an important property in common: the position of equilibrium is independent of the total pressure. The truth of this statement may be shown in a number of ways, one of which is the following. Suppose a mixture of a moles of hydrogen and b moles of iodine is heated until equilibrium is attained, when y moles of each have combined to form $2y$ moles of hydrogen iodide; the number of moles of hydrogen is then $a - y$ and of iodine $b - y$, but the total remains $a + b$.

The partial pressures of the reacting substances are

$$p_{\text{H}_2} = \frac{a-y}{a+b}P; \quad p_{\text{I}_2} = \frac{b-y}{a+b}P; \quad p_{\text{HI}} = \frac{2y}{a+b}P,$$

where P is the total pressure, so that the value of K_p is given by

$$K_p = \frac{p_{\text{HI}}^2}{p_{\text{H}_2} \times p_{\text{I}_2}} = \frac{\left(\frac{2y}{a+b}\right)^2 P^2}{\left(\frac{a-y}{a+b}\right)P \left(\frac{b-y}{a+b}\right)P} = \frac{4y^2}{(a-y)(b-y)}. \quad (66)$$

Since the total pressure does not appear in the final result, it is evident that y , the amount of each reactant decomposed at equilibrium, should be independent of the pressure. This has been verified experimentally, as the results given below, obtained by M. Bodenstein (1897), show.

Total Pressure	0.5	1.0	1.5	2.0 atm.
y	0.202	0.214	0.222	0.231

The small variations are probably to be attributed to experimental errors, deviations from the gas laws and adsorption of hydrogen iodide on the walls of the vessel.

The same conclusion, concerning the independence of the position of equilibrium of the total pressure, would be reached for any reaction in which equal numbers of molecules are involved in direct and reverse processes. The quantity P appears the same number of times in the numerator and denominator of the expression for the equilibrium constant and so cancels out. The fact that the position of equilibrium is not altered by change of pressure is also in harmony with the principle of mobile equilibrium. Since there is no variation in the number of molecules in either direction it is evident that the system is not able to adjust itself so as to counteract an increase or decrease of pressure; there should consequently be no change in the position of equilibrium.

By means of (66) an expression for the amount of hydrogen iodide ($2y$) formed at equilibrium can be derived; this may be done by solving the quadratic with the following result

$$2y = \frac{(a+b) - \sqrt{(a+b)^2 - 4ab(K-4)/K}}{(K-4)/K}, \quad (67)$$

where K is the equilibrium constant, which is independent of the concentration or pressure units. It will be noted that of the two solutions of the quadratic equation, only the one with the negative sign preceding the square root is possible.

Some of the experimental results obtained in a study of the hydrogen-iodine-hydrogen iodide equilibrium (A. H. Taylor and R. H. Crist, 1941) are given in Table 132, for a temperature of 457.6° C. The results in the

TABLE 132. THE $\text{H}_2 + \text{I}_2 \rightleftharpoons 2\text{HI}$ EQUILIBRIUM AT 457.6°C .

H_2 mole liter ⁻¹	I_2 mole liter ⁻¹	HI mole liter ⁻¹	K
5.617×10^{-3}	0.5936×10^{-3}	1.270×10^{-2}	48.38*
3.841	1.524	1.687	48.61*
4.580	0.9733	1.486	49.54*
1.696	1.696	1.181	48.48†
1.433	1.433	1.000	48.71†
4.213	4.213	2.943	48.81†

* Equilibrium approached from hydrogen and iodine.

† Equilibrium approached from hydrogen iodide.

last column are approximately constant, although equilibrium was approached from hydrogen and iodine alone in the first three cases, and from hydrogen iodide in the last three. The equilibrium constant for the reaction as written above is thus 48.7 at 457.6°C .

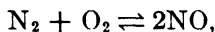
The values of K have been found to be 45.62 at 763.8°K . and 60.80 at 666.8°K .; if these results are inserted in (64), the increase in heat content ΔH for the reaction, over the temperature range considered, is found to be about $-3,000 \text{ cal.}$, i.e.,



the mean temperature being 715°K . From a knowledge of the heat capacities of the three reacting substances it would be possible to evaluate ΔH at 298°K ., but it is doubtful whether the available data are accurate enough for the results to have any significance. The method employed will, therefore, be illustrated later in another instance.

According to the Le Chatelier-Braun principle, increase of temperature should favor the reaction in which heat is absorbed. Since K for the hydrogen-iodine equilibrium decreases as the temperature is raised it is evident that increase of temperature inhibits the formation of hydrogen iodide from its elements. The reaction as written should therefore be exothermic; in other words ΔH should be negative, in agreement with the quantitative result obtained above from the van't Hoff equation.

The Formation of Nitric Oxide.—The combination of oxygen and nitrogen to form nitric oxide by passage through a flaming arc was once the chief method for the fixation of atmospheric nitrogen; the reversible reaction



which is involved, has been investigated by W. Nernst (1906) and others. The equilibrium constant

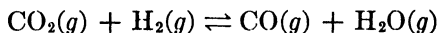
$$K = p_{\text{NO}}^2 / (p_{\text{N}_2} \times p_{\text{O}_2}) \quad (68)$$

increases with increasing temperature, the variation being represented by the equations

$$\ln K = -43,200/RT + 2.5 \quad \text{or} \quad \log K = -9554/T + 1.09 \quad (69)$$

over the range 1800° to 2700° K. By comparison with (63) it is seen that ΔH is 43,200 cal., so that the reaction is endothermic; this accounts for the increase in the yield of nitric oxide as the temperature is raised. With the aid of (68) the amount of nitric oxide formed at equilibrium, at different temperatures, for definite mixtures of nitrogen and oxygen can be calculated. The equilibrium constant is, of course, the same for partial pressures or concentrations.

The Water Gas Reaction.—The reversible reaction



is of considerable industrial importance and is generally known as the "water gas" reaction; it will be referred to frequently in this chapter from different points of view. The equilibrium composition, given by

$$K = \frac{p_{\text{CO}} \times p_{\text{H}_2\text{O}}}{p_{\text{CO}_2} \times p_{\text{H}_2}}, \quad (69)$$

using partial pressures for convenience, is independent of pressure; the process as written above is endothermic, i.e., ΔH is positive, and so K should increase with increasing temperature. For industrial purposes it is desirable to obtain a high yield of hydrogen from this reaction; it is necessary, therefore, to keep the temperature as low as possible. The use of excess of steam, making $p_{\text{H}_2\text{O}}$ high, will also tend to increase p_{H_2} , so that the expression for K may remain constant.

The molar heat capacities at constant pressure of the reacting substances, over a range of temperature, are given by the following expressions:

Final State	Initial State
$\text{CO} = 6.5 + 10^{-3}T$	$\text{H}_2 = 6.5 + 0.9 \times 10^{-3}T$
$\text{H}_2\text{O} = 8.81 - 1.9 \times 10^{-3}T$	$\text{CO}_2 = 7.0 + 7.1 \times 10^{-3}T$
$+ 2.22 \times 10^{-6}T^2$	$- 1.86 \times 10^{-6}T^2$
$\therefore \Delta C_P = 1.81 - 8.9 \times 10^{-3}T + 4.08 \times 10^{-6}T^2;$	
$\therefore \alpha = 1.81, \quad \beta = -8.9 \times 10^{-3}, \quad \gamma = 4.08 \times 10^{-6}.$	

The value of K has been found by experiment to be 0.534 at 959° K. and 1.571 at 1259° K., and substituting these figures, together with those for α , β and γ in (61),

$$\begin{aligned} \ln \frac{1.571}{0.534} = & -\frac{\Delta H_0}{R} \left(\frac{1}{1259} - \frac{1}{959} \right) + \frac{1.81}{R} \ln \frac{1259}{959} \\ & - \frac{4.45 \times 10^{-3}}{R} (300) + \frac{0.68 \times 10^{-6}}{R} [(1259)^2 - (959)^2] \end{aligned}$$

Inserting the value 1.987 cal. for R , this equation may be solved for ΔH_0 ,

which is found to be 10,100 cal. Since (see p. 214)

$$\Delta H = \Delta H_0 + \alpha T + \frac{1}{2}\beta T^2 + \frac{1}{3}\gamma T^3 + \dots, \quad (70)$$

it follows that

$$\Delta H = 10,100 + 1.81T - 4.45 \times 10^{-3}T^2 + 1.36 \times 10^{-6}T^3,$$

and so the change of heat content for the reaction may be calculated for any temperature in the range for which the heat capacity equations hold good.

Since ΔH_0 , α , β and γ are known, the integration constant in (62) may be calculated from the value of K at any one temperature; I/R is found to be +0.27, so that

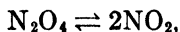
$$\begin{aligned} \ln K = & -\frac{10,100}{RT} + \frac{1.81}{R} \ln T - \frac{4.45 \times 10^{-3}}{R} T \\ & + \frac{0.68 \times 10^{-6}}{R} T^2 + 0.27; \end{aligned}$$

$$\begin{aligned} -\Delta F^0 = RT \ln K = & -10,100 + 1.81T \ln T - 4.45 \times 10^{-3}T^2 \\ & + 0.68 \times 10^{-6}T^3 + 0.54T. \end{aligned}$$

By means of these expressions the equilibrium constant at any temperature may be evaluated; the results are found to be in good agreement with experiment. These calculations show the use which may be made of the van't Hoff equation in conjunction with heat capacity data; the same methods may be applied to any reversible reaction. It must be borne in mind, however, that in the above reaction the equilibrium constant is the same whether partial pressures or concentrations are used; in other cases, where there is a change in the number of molecules, care must be taken to employ K_p in conjunction with heat capacities at constant pressure.²

Reactions With Change in the Number of Molecules.—When the numbers of molecules of reactants and products are different the *position* of equilibrium is affected by the total pressure, although theoretically the value of K_p should, of course, remain constant. It follows from the Le Chatelier principle that increase of pressure will tend to force the equilibrium in the direction in which there is a decrease in the number of molecules, and the same conclusion is reached from the law of chemical equilibrium. Although reactions of the type under consideration do not present any unusual features, a few examples will be considered so as to provide an opportunity for introducing certain points of interest.

The dissociation of nitrogen tetroxide is a familiar reaction; thus,



and the position of equilibrium under different conditions may be found by vapor density measurements, or better by observing the total pressure at constant volume for various temperatures (p. 318). Suppose the

system contains initially a moles of N_2O_4 per liter, and let y be the number of moles dissociated, leaving $a - y$ moles unchanged and producing $2y$ moles of NO_2 at equilibrium; the partial pressures of the two gases are thus

$$p_{\text{N}_2\text{O}_4} = \frac{a - y}{a + y} P \quad \text{and} \quad p_{\text{NO}_2} = \frac{2y}{a + y} P,$$

where P is the total pressure, and $a + y$ is the total number of moles at equilibrium. If the gases are assumed to be ideal, then the expression for the equilibrium constant is

$$K_p = \frac{p_{\text{NO}_2}^2}{p_{\text{N}_2\text{O}_4}} = \frac{\left(\frac{2y}{a + y}\right)^2 P^2}{\left(\frac{a - y}{a + y}\right) P} = \frac{4y^2 P}{a^2 - y^2}. \quad (71)$$

If the extent of dissociation is small, y^2 may be neglected in comparison with a^2 , and it follows from (71) that y/a , i.e., the fraction of the original N_2O_4 dissociated at equilibrium, is given by

$$y/a = \sqrt{K_p/4P}, \quad (72)$$

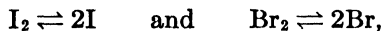
and so is approximately inversely proportional to the square root of the pressure. Increase of pressure will, therefore, decrease the degree of dissociation; that is, it shifts the equilibrium in the direction of nitrogen tetroxide in accordance with the principle of mobile equilibrium. Some of the experimental results for the reaction (F. H. Verhoek and F. Daniels, 1931) at 25°, 35° and 45° c. are given in Table 133. At each temperature

TABLE 133. THE $\text{N}_2\text{O}_4 \rightleftharpoons 2\text{NO}_2$ EQUILIBRIUM

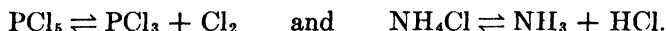
$a \times 10^3$ moles/liter	25°		35°		45°	
	P atm.	K_p	P atm.	K_p	P atm.	K_p
6.28	0.2118	0.1419	0.2382	0.3174	0.2662	0.6771
12.59	0.3942	0.1340	0.4396	0.2978	0.4896	0.6185
19.84	0.5996	0.1412	0.6623	0.3032	0.7349	0.6280
29.68	0.8623	0.1261	0.9470	0.2792	1.0474	0.5934

the values of K_p do not remain exactly constant; apart from individual variations there is a general decrease as the pressure is increased which may be attributed to departure of the gases from ideal behavior. When the data are extrapolated to zero pressure the true equilibrium constants are found to be 0.1426, 0.3183 and 0.6707 at 25°, 35° and 45° c. respectively. Assuming the heat of dissociation to remain constant over each 10° temperature range, the values of ΔH may be calculated by (64); these are 14,670 cal. for 25° to 35° c., and 14,530 cal. for 35° to 45° c. Since the dissociation reaction is endothermic, increase of temperature should favor dissociation, as it in fact does. The equilibrium constant increases as the temperature is raised, in agreement with the Le Chatelier principle.

Other dissociation reactions of a similar type are those in which the two atoms or molecules of product are identical, e.g.,



and others in which they are different, e.g.,



These reactions may all be studied by the density method, preferably by determining the pressure of a given mass at constant volume. In each case increase of pressure tends to suppress the dissociation, i.e., the position of equilibrium is driven to the left. For a small extent of decomposition the fraction of the initial material dissociated is in each case inversely proportional to the square root of the pressure, as shown by (72).³

Formation of Sulfur Trioxide.—The exothermic reversible reaction between sulfur dioxide and oxygen to produce the trioxide, viz., $2\text{SO}_2 + \text{O}_2 \rightleftharpoons 2\text{SO}_3$, is employed commercially in the manufacture of sulfuric acid, and the conditions of equilibrium have been studied fully. The flow method (p. 832) was employed by M. Bodenstein (1905) who obtained the data in Table 134; the experimental temperature was 1000°K . and the

TABLE 134. THE $2\text{SO}_2 + \text{O}_2 \rightleftharpoons 2\text{SO}_3$ EQUILIBRIUM AT 1000°K .

p_{SO_2}	p_{O_2}	p_{SO_3}	K_p
0.273 atm.	0.402 atm.	0.325 atm.	3.56
0.309	0.353	0.338	3.38
0.456	0.180	0.364	3.54
0.564	0.102	0.333	3.48

equilibrium constant was calculated from the partial pressures in atmospheres ($K_p = p_{\text{SO}_3}^2 / p_{\text{SO}_2}^2 \times p_{\text{O}_2}$). The variation of the equilibrium constant with temperature can be represented with fair accuracy, between 800° and 1170°K ., by the expression [cf. equation (63)]

$$\ln K_p = \frac{45,200}{RT} - 21.52 \quad \text{or} \quad \log K_p = \frac{9888}{T} - 9.346,$$

so that K_p decreases with increasing temperature, as expected for an exothermic process; the value of ΔH over the temperature range mentioned is assumed to be constant and is evidently $-45,200 \text{ cal}$. In order to obtain the maximum conversion of sulfur dioxide into trioxide the value of K_p should be large, and hence operations should be carried out at as low a temperature as possible. Since the reaction is slow under such conditions, a very long time being required for equilibrium to be attained, a catalyst is used to increase the velocity of the combination of sulfur dioxide and oxygen. A medium temperature, about 450°C ., is used as a compromise between the high reaction rate for high temperatures and the greater proportion of sulfur trioxide formed at equilibrium at low temperatures.

If the expression for K_p is written in terms of mole fractions of the reacting substances and the total pressure, it is

$$K_p = \frac{N \times n_{\text{SO}_2}^2}{n_{\text{SO}_2}^2 \times n_{\text{O}_2}} \cdot \frac{1}{P}, \quad (73)$$

where the n terms are the numbers of moles of the various species and N is the total number; hence,

$$\frac{n_{\text{SO}_2}}{n_{\text{SO}_2}} = \sqrt{K_p P \frac{n_{\text{O}_2}}{N}}, \quad (74)$$

so that the ratio of trioxide to dioxide at equilibrium is proportional to the square root of the pressure, if n_{O_2}/N is constant. Hence increase of pressure will increase the yield of trioxide, but the effect is so relatively small that no advantage is taken of it in industrial practice. It will be seen from (74) that if other conditions remain unchanged, and the total number of moles N is increased by the presence of an inert gas, e.g., nitrogen, the yield of sulfur trioxide is diminished.

The Combination of Nitrogen and Hydrogen.—The exothermic formation of ammonia from nitrogen and hydrogen, viz., $\text{N}_2 + 3\text{H}_2 \rightleftharpoons 2\text{NH}_3$, was investigated thoroughly by F. Haber (1905–15) and more recently by A. T. Larson and R. L. Dodge (1923–24) and others, because the reaction is of such great industrial significance. The maximum conversion of nitrogen and hydrogen into ammonia will be obtained when the reactants are present in their stoichiometric ratio, i.e., one mole of nitrogen to three of hydrogen, and much of the experimental work has been carried out with such mixtures. Suppose there are present initially 1 mole of nitrogen and 3 moles of hydrogen and y moles of the former are converted into ammonia at equilibrium; there are then present $1 - y$ moles of nitrogen, $3(1 - y)$ moles of hydrogen, and $2y$ moles of ammonia, the total being $4 - 2y$ moles. The expression for K_p is then

$$K_p = \frac{p_{\text{NH}_3}^2}{p_{\text{N}_2} \times p_{\text{H}_2}^3} = \frac{\left(\frac{2y}{4 - 2y}\right)^2 P^2}{\left(\frac{1 - y}{4 - 2y}\right) \left(\frac{3 - 3y}{4 - 2y}\right)^3 P^4} = \frac{16y^2(2 - y)^2}{27P^2(1 - y)^4}, \quad (75)$$

and if y is small in comparison with unity this reduces to

$$K_p = \frac{64y^2}{27P^2} \quad \text{or} \quad y^2 = \frac{27}{64} K_p P^2, \quad (76)$$

when y , the fraction of the nitrogen converted into ammonia, is seen to be directly proportional to the total pressure P . The result is in qualitative agreement with that to be expected from the Le Chatelier-Braun principle. Increase of pressure should tend to move the position of equilibrium in the direction of a decrease in the number of molecules, that is, it should favor the formation of ammonia.

TABLE 135. THE $\text{N}_2 + 3\text{H}_2 \rightleftharpoons 2\text{NH}_3$ EQUILIBRIUM

Tempera- ture	Total Pressure in atm.					
	10	50	100	300	600	1000
	Percentage of Ammonia at Equilibrium					
623° K.	10.38	25.11	—	—	—	—
673	3.85	15.11	24.91	—	—	—
723	2.04	9.17	16.36	35.5	53.6	69.4
773	1.20	5.58	10.40	26.2	42.1	—
	Equilibrium Constant: K_p *					
623° K.	7.08×10^{-4}	7.73×10^{-4}	—	—	—	—
673	1.66	1.69	1.88×10^{-4}	—	—	—
723	0.434	0.476	0.526	0.782×10^{-4}	1.67×10^{-4}	5.43×10^{-4}
773	0.145	0.150	0.162	0.248	0.424	—

Of the results quoted in Table 135 (A. T. Larson and R. L. Dodge, 1923-24), the upper portion gives the percentage of ammonia formed at equilibrium, from a 1 to 3 mixture of nitrogen and hydrogen, at various temperatures and total pressures, whereas in the lower the corresponding equilibrium constants with partial pressures in atm. are recorded. From the table it will be seen that at 773° K., the amount of ammonia formed is relatively small and is approximately proportional to the pressure at the lower pressures, in agreement with (76). The relationship no longer holds when the conversion into ammonia occurs to a large extent, but the results show clearly that increase of pressure markedly increases the yield of ammonia. In the commercial operation of the Haber process the gas pressure is about 200 atm., and in certain modifications pressures as high as 1000 atm. have been employed. The values of K_p in Table 135 are remarkably constant up to pressures of 100 atm., thus providing striking confirmation of the law of equilibrium; at higher pressures considerable variations are observed which are undoubtedly due to deviations from ideal gas behavior under these conditions. This has been proved by using activities or fugacities, derived from P - V - T data (cf. p. 302), in place of pressures in the equilibrium constant expression; in this way the constancy has been extended to pressures as high as 600 atm. or more.⁴

Increase of temperature is seen to result in a decrease in the yield of ammonia, as is to be anticipated since the reaction is exothermic. In order to produce the maximum conversion of nitrogen into ammonia it is desirable, therefore, to work at as low a temperature as is compatible with an appreciable rate of reaction. Catalysts are employed in order to expedite the process, and the operating temperature is about 500° C. From the heat capacities of nitrogen, hydrogen and ammonia (p. 215) it is seen that

$$\Delta C_P = -9.92 - 2.3 \times 10^{-3}T + 10.2 \times 10^{-6}T^2;$$

$$\therefore \alpha = -9.92, \quad \beta = -2.3 \times 10^{-3} \quad \text{and} \quad \gamma = 10.2 \times 10^{-6}.$$

* The values of K_p in the original publication are based on the equation $\frac{1}{2}\text{N}_2 + \frac{3}{2}\text{H}_2 \rightleftharpoons \text{NH}_3$, and hence are equal to the square root of the results given here.

Using these values of α , β and γ , and those of K_p from Table 135, the following relationship has been derived (cf. p. 830) to represent the variation of K_p with temperature between 600° and 800° K. for pressures up to 100 atm.:

$$\ln K_p = \frac{18,550}{RT} - \frac{9.92}{R} \ln T - \frac{1.15 \times 10^{-3}}{R} T + \frac{1.7 \times 10^{-6}}{R} T^2 + 19.72$$

or

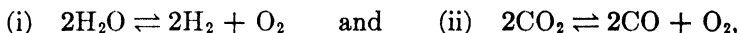
$$-\Delta F^0 = RT \ln K_p = 18,550 - 9.92T \ln T - 1.15 \times 10^{-3} T^2 + 1.7 \times 10^{-6} T^3 + 39.18T.$$

It follows, therefore, that over the same temperature range

$$\Delta H = -18,550 - 9.92T - 1.15 \times 10^{-3} T^2 + 3.4 \times 10^{-6} T^3,$$

and so the change of heat content may be calculated at any temperature. If the relationship applies at 298° K., i.e., 25° C., the heat of formation of one mole of ammonia is -10,675 cal.

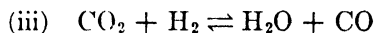
Combination of Equilibria.—For the thermal dissociation of (i) steam and (ii) carbon dioxide,



the equilibrium constants, involving partial pressures, are

$$K_1 = \frac{p_{H_2}^2 \times p_{O_2}}{p_{H_2O}^2} \quad \text{and} \quad K_2 = \frac{p_{CO}^2 \times p_{O_2}}{p_{CO_2}^2}. \quad (77)$$

In a mixture of steam and carbon dioxide both equilibria will coexist, and the partial pressures of the various constituents will be adjusted so that the equilibrium constants in the mixture are the same as in the separate systems. Simultaneously with the two processes mentioned, the water gas reaction



must also attain equilibrium, and for this

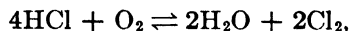
$$K_3 = \frac{p_{H_2O} \times p_{CO}}{p_{CO_2} \times p_{H_2}} = \sqrt{\frac{K_2}{K_1}}, \quad (78)$$

so that K_3 is related to K_1 and K_2 . This relationship may be tested by experimental data, since the three equilibria have been studied separately, and the results are recorded in Table 136. In view of the very small

TABLE 136. COMBINATION OF EQUILIBRIA

temperature	K_1	K_2	$\sqrt{K_2/K_1}$	K_3
1395° K.	2.1×10^{-13}	1.4×10^{-12}	2.6	2.0
1565	1.6×10^{-11}	1.3×10^{-10}	2.9	2.8
1823	7.0×10^{-9}	3.2×10^{-8}	2.2	3.5

magnitudes of K_1 and K_2 , which have been obtained by extrapolation, the agreement between $\sqrt{K_2/K_1}$ and K_3 may be regarded as satisfactory; the value of the former at 1823° K. is undoubtedly too low, since K_3 increases with increasing temperature. In this particular case the investigation of the equilibrium for the water gas reaction as a whole can be carried out more accurately than that of the partial reactions; the same holds good for the reactions involved in the Deacon process for the production of chlorine, viz.,



for which the partial equilibria are



In other instances, however, the situation may be reversed, and the method described may be used to evaluate equilibrium constants not easily determined experimentally. Sometimes the complete reaction and one of the partial reactions can be studied; in that case the constant for the other partial equilibrium may be calculated.

Equilibria in Liquid Systems.—One of the most familiar instances of a reversible reaction in a homogeneous liquid system is that between ethyl alcohol and acetic acid (cf. p. 816). If the system were ideal, either Henry's law or Raoult's law being applicable to the reactants and products, then the equilibrium constant would be

$$K_x = \frac{x_{\text{CH}_3\text{CO}_2\text{C}_2\text{H}_5} \times x_{\text{H}_2\text{O}}}{x_{\text{C}_2\text{H}_5\text{CO}_2\text{H}} \times x_{\text{C}_2\text{H}_5\text{OH}}} = \frac{n_{\text{CH}_3\text{CO}_2\text{C}_2\text{H}_5} \times n_{\text{H}_2\text{O}}}{n_{\text{C}_2\text{H}_5\text{CO}_2\text{H}} \times n_{\text{C}_2\text{H}_5\text{OH}}}, \quad (79)$$

where x and n are the mole fractions and moles respectively of the several reacting species; * the total number of moles does not appear in the final fraction because there are two terms in both numerator and denominator. Berthelot and St. Gilles took known amounts of acid and alcohol, and sealed them in glass tubes which were heated to over 100° C. to expedite the attainment of equilibrium; the tubes were then rapidly cooled and the contents analyzed. If a and b are the initial numbers of moles of acid and alcohol, respectively, and y is the number of moles of ester formed at equilibrium, then an equal amount of water is produced, and $a - y$ and $b - y$ moles of acid and alcohol, respectively, remain. It follows, therefore, that

$$K = \frac{y \times y}{(a - y)(b - y)} \quad (80)$$

and

$$y = \frac{(a + b) - \sqrt{(a + b)^2 - 4ab(K - 1)/K}}{2(K - 1)/K}. \quad (81)$$

* In many elementary textbooks the equilibrium constant for this process is given in terms of concentrations expressed in moles per liter; although this gives the same value of K as in (79), because the number of molecules in the reaction is unchanged, the procedure is misleading and should be avoided.

TABLE 137. THE ACETIC ACID-ETHYL ALCOHOL-ETHYL ACETATE-WATER EQUILIBRIUM

Initial Moles Acetic Acid	Initial Moles Ethyl Alcohol	Equilibrium Moles Ethyl Acetate	K_z	y (calc.)
1.0	0.18	0.171	3.9	0.171
1.0	0.33	0.293	3.3	0.301
1.0	0.50	0.414	3.4	0.423
1.0	1.00	0.667	4.0	0.667
1.0	2.00	0.858	4.6	0.850
1.0	8.00	0.966	3.9	0.970

The equilibrium constants are recorded in Table 137, together with the values of y calculated on the assumption that K is equal to 4, at the experimental temperature. The figures in the fourth column are not remarkably constant, but the deviations are probably due, to a great extent, to experimental errors; this may be seen by comparing the observed (third column) and calculated (fifth column) amounts of ester formed, the latter being based on an assumed constant value for K_z . The agreement is seen to be good, so that only very small changes in the observed y values would have given a more consistent equilibrium constant. It must also be remembered that departure from ideal behavior will also be a source of error.

The equilibrium constant for the reaction is almost independent of temperature. It follows from the reaction isotherm, or the Le Chatelier principle, that the heat of reaction should be zero or very small; this is in agreement with experiment.

It is of interest to note that the reaction has also been studied in the vapor phase, and the equilibrium constant found to be different from the value for the liquid system. By imagining the liquid and vapor systems in equilibrium, it can be readily seen that the partial pressures of the reacting species in the vapor are equal to the vapor pressures of the various substances as present in the liquid phase. If the two systems are ideal, then $p_A = x_A p_A^0$ (Raoult's law), where p_A is the partial pressure of the reactant A in the vapor, x_A is its mole fraction in the liquid, and p_A^0 the vapor pressure of the pure liquid; it follows, therefore, since K_p and K_z are identical in the vapor phase, that

$$(K_z)_{\text{vapor}} = (K_z)_{\text{liquid}} \times \frac{p_{\text{C}_2\text{H}_5\text{CO}_2\text{C}_2\text{H}_5}^0 \times p_{\text{H}_2\text{O}}^0}{p_{\text{C}_2\text{H}_5\text{CO}_2\text{H}}^0 \times p_{\text{C}_2\text{H}_5\text{OH}}^0}, \quad (82)$$

the p^0 values being measured at the experimental temperature.⁵

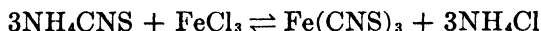
Dissociation of Nitrogen Tetroxide in Solution.—The dissociation of nitrogen tetroxide (cf. p. 836) has also been studied in chloroform solution (J. T. Cundall, 1891); the extent of dissociation was determined by the darkening in color of the liquid. The data in Table 138 have been calculated from Cundall's results; since the solutions are dilute the use of molar concentrations is justifiable in the equilibrium constant (p. 822), and so

TABLE 138. DISSOCIATION OF NITROGEN TETROXIDE IN CHLOROFORM SOLUTION AT 8.2° C.

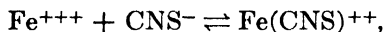
$c_{N_2O_4}$	c_{NO_2}	K_c
0.129 mole/liter	1.17×10^{-3} mole/liter	1.07×10^{-6}
0.227	1.61	1.14
0.324	1.85	1.05
0.405	2.13	1.13
0.778	2.84	1.04

K_c , given by $c_{NO_2}^2/c_{N_2O_4}$ at equilibrium, is recorded. The constancy of the figures in the last column provides a satisfactory verification of the approximate form of the equilibrium expression. It is evident that in the dilute solution studied the concentrations of the reacting substances are not very different from their activities.

The Ferric-Thiocyanate Reactions.—The reaction



is frequently quoted as an example of a reversible reaction occurring in solution; the addition of excess of ammonium thiocyanate or of ferric chloride results in an increase in intensity of the red color, attributed to ferric thiocyanate, whereas ammonium chloride causes the color to become paler. It is now believed, however, that the red color is produced by the complex anion $Fe(CNS)^{++}$, and that the equilibrium is ionic (see p. 588); thus,



so that excess of ferric or thiocyanate ions results in an increased formation of the complex. The addition of ammonium chloride changes the activities to such an extent that the complex ion tends to dissociate, and other neutral electrolytes have been shown to exert a similar effect.⁶

Effect of Solvent on Equilibrium.—When an equilibrium is established in a solvent the system is virtually equivalent to a gaseous equilibrium to which an inert gas has been added at constant pressure (cf. p. 823). If the reaction involves no change in the number of molecules, as in the esterification reaction described above, the solvent should have no influence on the position of equilibrium; this has been found to be true in carbon tetrachloride solution. In many cases, however, the solvent has a considerable effect on the activities of the reacting substances, and hence the value of the equilibrium constant may differ from one solvent to another. The matter was first considered by J. H. van't Hoff (1898), who showed that allowance could be made for the change in active mass, which is equivalent to what is now called activity, by taking into consideration the solubilities in the particular solvent of the substances concerned in the reaction. Since a saturated solution is in equilibrium with solid, the chemical potential of the solute in the former will be equal to that of the solid. If the latter is taken as the standard state, that is the state of unit activity, then the solute in the saturated solution will also have unit activity. The activity of the solute in any solution of concentration c will then be c/s , where s is the saturation solubility in the given solvent. For a simple reversible reaction $A \rightleftharpoons B$, the equilibrium constant is $K = a_B/a_A$, and express-

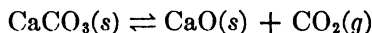
ing the activities in terms of the solubilities, it follows that

$$K = \frac{c_B/s_B}{c_A/s_A},$$

which should be independent of the nature of the solvent, since the same standard states, viz., the pure solids, are chosen in each case. The isomeric change of a triazole to a diazoimide was studied in a number of media by O. Dimroth (1910); he found that while the concentration equilibrium constant c_B/c_A varied from 0.32 to 21.7, the activity constant, as given above, changed only from 0.32 to 0.4. It may be remarked that strictly speaking the c and s terms should be in mole fractions, and it is only for dilute solutions that the approximation made by Dimroth of using moles per liter is justifiable.⁷

HETEROGENEOUS CHEMICAL EQUILIBRIA

Chemical Equilibria in Heterogeneous Systems.—In their development of the law of mass action, Guldberg and Waage concluded, from empirical considerations, that whenever a solid substance was involved in a reversible reaction its "active mass" should be regarded as constant, irrespective of the amount present. This means that the expression for the equilibrium constant need contain no terms for any substance present as solid at equilibrium. For example, suppose calcium carbonate is heated in a closed space; it undergoes dissociation, and the equilibrium



is established. The calcium carbonate and oxide are present as solids, and so their "active masses" are constant; hence the equilibrium constant will depend only on the partial pressure of the carbon dioxide, that is

$$K_p = p_{\text{CO}_2}. \quad (83)$$

Since K_p is a constant for a given temperature, it follows that at every temperature there should be a definite pressure of carbon dioxide, i.e., the dissociation pressure (p. 779), in equilibrium with calcium carbonate and oxide. This conclusion is in harmony with the phase rule and with experiment.

In the language of thermodynamics, the chemical potential and activity of a pure solid are constant at constant temperature and external pressure. Consider a pure solid in equilibrium with its vapor; the chemical potential of the latter is $\mu^0 + RT \ln p$, where p is the pressure (fugacity) of the vapor. Since the solid and vapor are in equilibrium, this will also represent the chemical potential of the former. At constant temperature and external pressure, p is constant, and hence under these conditions the chemical potential of the pure solid will be constant. Just as for a liquid solution [cf. equation (86), p. 684], the chemical potential of a component of a solid solution may be represented by $\mu = \mu_x^0 + RT \ln a$, where μ_x^0 is constant. As seen above, for a pure solid μ is constant, and hence it is evident that the activity must also be con-

stant. If the pure solid is chosen as the standard state, just as is a pure liquid, then the constant activity is set equal to unity. It is clear, therefore, that this is equivalent to the statement made earlier, that the terms for substances present as pure solids may be omitted from the expression for the equilibrium constant.

The same conclusion may be reached from (16), which gives the general thermodynamic condition for any chemical equilibrium. Since the chemical potentials of solid phases remain constant, the inclusion of solid reactants and products will only involve the addition of constant terms to either or both sides of the equation. The expression for the equilibrium constant will consequently contain factors, e.g., partial pressures or, in general, activities, only for those species present in the gaseous state. The same general rule concerning the omission of the activities, or partial pressures, of solid reactants and products applies, of course, in connection with the reaction isotherm, e.g., (37) and (44).

Strictly speaking, the vapor pressure and hence the chemical potential of a solid are constant only if the total pressure, as well as the temperature, is fixed. It follows, therefore, that the omission of terms for solid substances is only justifiable provided the pressure of the system is unchanged. Unless extremely high pressures are involved, however, the change in vapor pressure is negligible and so the slight error incurred may be disregarded; the equilibrium constant, with terms for gaseous substances only, can thus be taken as being independent of the total pressure.

Attention should be called to the fact that although the van't Hoff equation (53) or (54), and its various integrated forms, are applicable to heterogeneous reactions, the value of ΔH is for the complete reaction *including solid phases*. This will differ from the corresponding ΔH for the reaction taking place entirely in the gas phase by the heats of sublimation of the solids.

From the foregoing discussion it will be apparent that the basis of the constancy of the chemical potential and activity (or "active mass") of a solid lies in its constant vapor pressure. An alternative physical, rather than thermodynamic, point of view is therefore to consider a heterogeneous reaction as taking place homogeneously in the gas phase. For example, in the dissociation of calcium carbonate, the equilibrium constant of the *homogeneous* gas reaction is

$$K'_p = p_{\text{CaO}} \times p_{\text{CO}_2} / p_{\text{CaCO}_3}, \quad (84)$$

where p_{CaCO_3} and p_{CaO} are the respective partial pressures in the gas phase. Since the calcium carbonate and oxide vapors are in equilibrium with their respective solids, the values of p_{CaCO_3} and p_{CaO} are equal to the vapor pressures at the experimental temperatures; these are constants, independent of the quantity of solid present, and so it follows that

$$p_{\text{CO}_2} = \text{constant},$$

as found above.

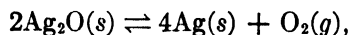
Solid-Gas Reactions: Two Solids and one Gas.—Instances of the application of the equilibrium constant and related matters will be considered for various types of reaction, the number of solids and gases concerned being used as a means of classification. The dissociation of a metallic carbonate, e.g.,



of a higher to a lower oxide, e.g.,



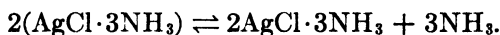
and of an oxide to a metal, e.g.,



are examples of equilibria of two solids and a gas. Other illustrations are the dissociation of a salt hydrate, e.g.,



or of a metallic ammine, e.g.,



In every case, when equilibrium is attained in a closed space, there will be a definite dissociation pressure of gas at each temperature; as seen above, this pressure represents the appropriate value of K_p for the whole *system*. The word system is emphasized, since the definite pressure is only obtained if both solid phases are present at equilibrium (cf. p. 783).

If the dissociation pressure is known at two temperatures the heat of dissociation over this range of temperature may be evaluated by the integrated form of the van't Hoff equation, i.e., (64). As a general rule insufficient data are available for the use of the accurate equation (61) which allows for the variation of ΔH with temperature, and so the simple form is used. For the present case it may be written

$$\ln \frac{p_2}{p_1} = -\frac{\Delta H}{R} \left(\frac{1}{T_2} - \frac{1}{T_1} \right) \quad \text{or} \quad \log \frac{p_2}{p_1} = -\frac{\Delta H}{2.30R} \left(\frac{1}{T_2} - \frac{1}{T_1} \right), \quad (85)$$

where p_1 and p_2 are the dissociation pressures at temperatures T_1 and T_2 . The dissociation of calcium carbonate has been frequently studied, and the results in Table 139 give the equilibrium pressures of carbon dioxide

TABLE 139. DISSOCIATION OF CALCIUM CARBONATE

Temperature	Dissociation Pressure	ΔH
773° K.	0.773 mm.	43.2 kcal.
873	1.84	42.1
973	22.2	41.8
1073	167	39.0
1173	793	88.9
1273	2942	

at a number of temperatures (F. H. Smyth and L. H. Adams, 1923). The values of ΔH for each range of 100° are included; these may be compared with the thermochemical value of 42.9 kcal. at room temperature. Since ΔH is positive the reaction is endothermic and, in agreement with the Le Chatelier principle, increase of temperature favors dissociation.⁸

The dissociation of other carbonates has been investigated and an empirical rule has been found connecting the heat of dissociation and the dissociation pressure (cf. p. 864): in general, the greater the heat absorbed on dissociation the lower the pressure at a given temperature. In the alkaline-earth carbonate series, for example, the heats of dissociation increase in the order $\text{Ca} < \text{Ba} < \text{Sr}$, and the temperatures at which dissociation pressures of 1 atm. of carbon dioxide are attained increase in the same order.

One Solid and Two Gases.—The classical examples of this type of heterogeneous equilibrium are the dissociation of ammonium salts, first studied quantitatively by N. F. Isambert (1881–83); for ammonium hydrosulfide, for instance,



it follows that

$$K_p = p_{\text{NH}_3} \times p_{\text{H}_2\text{S}}. \quad (86)$$

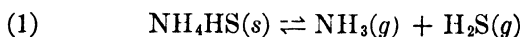
Since equivalent amounts of ammonia and hydrogen sulfide are produced, p_{NH_3} and $p_{\text{H}_2\text{S}}$ will be equal, provided the space into which dissociation occurs does not contain initially an excess of either gas; the total pressure P is equal to the sum of p_{NH_3} and $p_{\text{H}_2\text{S}}$, and hence each will be equal to $P/2$, and K_p will be $P^2/4$. If the space above the solid contains excess of either of the gaseous products, a different situation arises. Let p be the initial pressure of one of the gases, e.g., ammonia, and let y be the pressure of each gas produced by dissociation, which must of course be the same for both; then p_{NH_3} is equal to $p + y$, whereas $p_{\text{H}_2\text{S}}$ is equal to y , and so K_p is $(p + y)y$, and the total pressure is $p + 2y$. If p is known and the final total pressure is measured, both y and K_p can be evaluated. Many experiments of this type were made by Isambert on ammonium hydrosulfide and ammonium cyanide, and the results were found to be in satisfactory agreement with the requirements of the law of equilibrium for a heterogeneous reaction; some of the data for the former compound are reproduced in Table 140, the partial pressures being expressed in atm. The figures in the last column show a satisfactory constancy for K_p , either

TABLE 140. DISSOCIATION OF AMMONIUM HYDROSULFIDE AT 24°C .

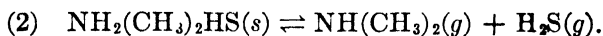
p_{NH_3}	$p_{\text{H}_2\text{S}}$	K_p
0.309 atm.	0.309 atm.	9.56×10^{-2}
0.364	0.258	9.40
0.539	0.174	9.37
0.249	0.369	9.16
0.161	0.595	9.55

when dissociation occurs in the absence of excess of the products, as in the first line, or when there is excess of ammonia, as in the second and third, or excess of hydrogen sulfide, as in the last two lines. Somewhat similar results were obtained with ammonium cyanide.

An interesting application of the law of chemical equilibrium was made by J. Walker and J. S. Lumsden (1897) when investigating the dissociation of a mixture of two hydrosulfides, e.g.,



and

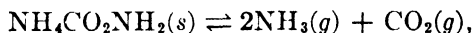


If p_1 and p_2 are the partial pressures of ammonia and of dimethylamine in the gas mixture, then the partial pressure of hydrogen sulfide must be the sum of these two, i.e., $p_1 + p_2$; it follows, therefore, that

$$p_1(p_1 + p_2) = K_1 \quad \text{and} \quad p_2(p_1 + p_2) = K_2, \quad (87)$$

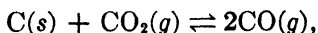
where K_1 and K_2 are the two equilibrium constants. If these are known, from experiments on the two hydrosulfides separately, then p_1 and p_2 may be calculated and the total gas pressure, $2p_1 + 2p_2$, evaluated and compared with the observed value. The comparison was made for a number of pairs of hydrosulfides at several temperatures, and general agreement was found.

The dissociation of ammonium carbamate is unusual for an ammonium salt, for three molecules of gas are produced; thus,



so that $K_p = p_{\text{NH}_3}^2 \times p_{\text{CO}_2}$. If P is the total pressure of the gases at any temperature, then p_{NH_3} must be $2P/3$ whereas p_{CO_2} is $P/3$; hence K_p in terms of the total pressure is $4P^3/27$. Since the equilibrium constant involves $p_{\text{NH}_3}^2$, it is evident that excess of ammonia should repress the dissociation of ammonium carbamate to a greater extent than a similar excess of carbon dioxide; this expectation has been confirmed by experiment.⁹

A different kind of reaction falling into the same category, viz., one solid and two gases, is the technically important equilibrium



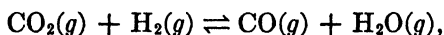
for which $K_p = p_{\text{CO}}^2/p_{\text{CO}_2}$. The results of different workers do not always coincide, but according to International Critical Tables the equilibrium constant between 1100° and 1500° K. is best represented by the expression

$$RT \ln K_p = -40,900 + 4.9T \ln T - 4.95 \times 10^{-2}T^2 + 5.1 \times 10^{-7}T^3 + 12.66T,$$

based on the following data for the heat capacities:

$$\begin{aligned} 2\text{CO}(g): \quad C_P &= 13.0 + 2.0 \times 10^{-3}T; \\ \text{C(graphite)}: \quad C_P &= 1.1 + 4.8 \times 10^{-3}T + 1.2 \times 10^{-6}T^2, \\ \text{CO}_2(g): \quad C_P &= 7.0 + 7.1 \times 10^{-3}T + 1.86 \times 10^{-6}T^2, \\ \therefore \Delta C_P &= 4.9 - 9.9 \times 10^{-3}T + 3.06 \times 10^{-6}T^2. \end{aligned}$$

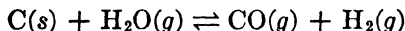
For the homogeneous water gas reaction



the equilibrium constant, as seen on p. 836, is given by

$$RT \ln K_p = -10,100 + 1.81T \ln T - 4.45 \times 10^{-3}T^2 + 6.8 \times 10^{-7}T^3 + 0.54T.$$

Subtraction of the two equations gives the variation with temperature of the equilibrium constant of the reaction

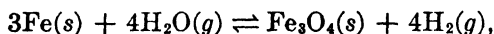


as

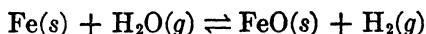
$$RT \ln K_p = -39,800 + 2.89T \ln T - 0.5 \times 10^{-3}T^2 - 1.7 \times 10^{-7}T^3 + 12.12T,$$

where K_p is $p_{\text{CO}} \times p_{\text{H}_2}/p_{\text{H}_2\text{O}}$. This is an interesting illustration of the combination of a heterogeneous and a homogeneous equilibrium, and many others are to be found in the chemical literature.

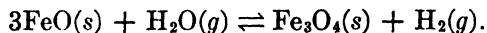
Two Solids and Two Gases.—The reaction between steam and iron provides an instance of this type of system; the reversible equilibrium is often represented by the equation



but at temperatures above 850° K. the oxide FeO is also formed, and two additional equilibria are involved, viz.,



and



An examination of these equations shows that in each case the law of chemical equilibrium leads to the result

$$p_{\text{H}_2}/p_{\text{H}_2\text{O}} = \text{constant}, \quad (88)$$

which is in agreement with experiment, as is shown by the data (G. Preuner, 1904) in Table 141 for 1173° K. Analogous processes to those described involve carbon dioxide, instead of water, the reaction product being carbon monoxide. By combining the equilibrium

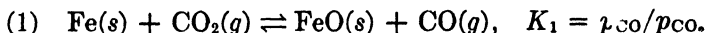
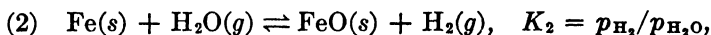


TABLE 141. THE IRON-STEAM EQUILIBRIUM AT 1173° K.

p_{H_2}	p_{H_2O}	p_{H_2}/p_{H_2O}
13.5 mm.	8.8 mm.	1.53
26.0	17.4	1.49
54.1	35.4	1.53
71.8	49.3	1.46

with that for the corresponding reaction



it is seen that $K_1/K_2 = p_{CO} \times p_{H_2O}/p_{CO_2} \times p_{H_2}$, which is the equilibrium constant (K_3) for the water gas reaction. The experimental data for K_1 and K_2 , and K_1/K_2 calculated from them, are given in Table 142,

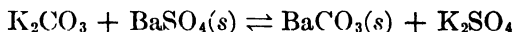
TABLE 142. COMBINATION OF HETEROGENEOUS EQUILIBRIA

Temperature	973°	1073°	1173°	1273°
K_1	1.47	1.81	2.15	2.48
K_2	2.38	2.00	1.67	1.49
K_1/K_2	0.65	0.91	1.29	1.66
$K_3(\text{obs.})$	0.63	0.93	1.27	1.67

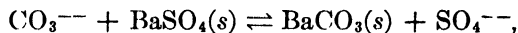
together with the observed values for K_3 . Equilibria of a similar type, using cobalt or tin instead of iron, have been studied and the constant for the water gas equilibrium derived in an analogous manner (P. H. Emmett, 1933).¹⁰

Solid-Liquid Equilibria.—The chemical potential and activity of a solid in contact with a liquid phase are also constant at definite temperature and pressure. As with solid-gas equilibria, the standard state is chosen as the pure solid, and so the activity of the latter is taken as unity. This means that as in the case of heterogeneous gas reactions, terms involving solids may be omitted from the expressions for the equilibrium constant or free energy change for solid-liquid reactions.

The reaction between potassium carbonate solution and solid barium sulfate, or between potassium sulfate solution and solid barium carbonate, viz.,



or, expressed ionically,



were studied by C. M. Guldberg and P. Waage (1879). Taking the activities of the solids as unity, and assuming the solutions to be so dilute that concentrations, in g. ions per liter, may be used for the activities of the ions, the equilibrium constant is given by

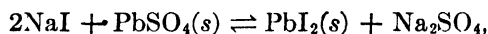
$$K_c = c_{\text{SO}_4^{--}}/c_{\text{CO}_3^{--}}. \quad (89)$$

To test this equation solid barium sulfate was allowed to come to equilibrium with solutions of potassium carbonate, some of which contained

TABLE 143. THE BARIUM SULFATE-CARBONATE EQUILIBRIUM IN SOLUTION

Initial Concentrations		Equilibrium Concentrations		$c_{\text{SO}_4^{--}}/c_{\text{CO}_3^{--}}$
Carbonate	Sulfate	Carbonate	Sulfate	
0.200	—	0.0395	0.1605	4.07
0.250	—	0.050	0.200	4.00
0.350	—	0.072	0.278	3.86
0.250	0.025	0.055	0.220	4.00
0.300	0.025	0.066	0.259	3.93
0.200	0.05	0.050	0.200	4.00

added potassium sulfate, and the final liquid was analyzed for sulfate and carbonate in solution. The results obtained are reproduced in Table 143; concentrations are in g. equiv. per liter, but the equilibrium constant is obviously independent of the concentration units. It must be admitted that the approximate constancy of the figures in the last column is better than might have been expected, in view of the fact that electrolyte solutions containing bivalent ions at the concentrations employed do not behave ideally; there may be, however, a cancellation of errors. The somewhat similar reaction



for which the equilibrium constant should be given by $c_{\text{SO}_4^{--}}/c_{\text{I}^-}^2$, was studied by A. Findlay (1900), but the "constant" was found to vary with the total ionic concentration. If the observations had been made in very dilute solutions it is probable that more satisfactory results would have been obtained. There is no doubt that the activity of a solid is to be regarded as constant, but systems involving sparingly soluble salts are better considered from the standpoint of the solubility product, allowance being made for the departure from ideal behavior; this subject will be considered in Chapter XII.

FREE ENERGY CHANGE

Affinity and Free Energy.—At the beginning of this chapter it was shown how the old ideas of affinity led to the development of the law of mass action, which emphasized the importance of "active mass" in determining the direction of chemical change. It was mentioned, too, that the original view of Guldberg and Waage that the equilibrium constant represented the ratio of the affinities of direct and reverse reactions was subsequently discarded. An alternative concept of affinity was proposed by J. Thomsen (1853) and M. Berthelot (1867), as indicated on p. 211: it was thought that the direction of chemical change was related to the heat evolved, but consideration of a reversible reaction shows immediately that the suggestion is untenable. It is to J. H. van't Hoff (1883) that credit must be given for the modern view, although the same idea was in the minds of J. Willard Gibbs (1876) and of H. von Helmholtz (1882). Van't Hoff proposed that affinity should be measured by the

quantity that is now known as the decrease of free energy accompanying the reaction. This is in harmony with the conclusion, derived from (134a), p. 234, that at constant temperature and pressure a process, in a closed system, can take place spontaneously in the direction of a decrease of free energy only. If the value of ΔF under a given set of conditions is positive, that is to say, there is an increase of free energy, the reaction cannot occur under those conditions. It may, nevertheless, be able to take place if the conditions, such as temperature and partial pressures, or concentrations, of the reactants and products, are changed appropriately. This may be readily seen from the general equation (44) for the free energy change in any reaction with arbitrary activities of reactants and products. If the process is to take place spontaneously, ΔF must be negative, so that from (44)

$$\ln K > \sum \nu \ln a. \quad (90)$$

In order to achieve this condition the equilibrium constant K may be increased by altering the temperature, or $\sum \nu \ln a$ may be decreased by decreasing the activities, i.e., concentrations or partial pressures, of the products or increasing those of the reactants, or both. It should be noted that although a decrease of free energy is a necessary requisite for a spontaneous reaction, there is no certainty that the process will take place at an appreciable speed. The factors which determine the *rate* of a chemical reaction, as distinct from the possibility of the reaction occurring, will be considered in Chapter XIII.

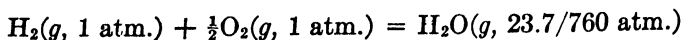
Determination of Free Energy Changes.—In view of the significance of the decrease of free energy as the criterion of spontaneous reaction, it is desirable that methods should be available for the determination of free energy changes. In this connection the reaction isotherm (44) which, in its most general form, is

$$-\Delta F = RT \ln K - RT \sum \nu \ln a, \quad (91)$$

is of fundamental importance, since it gives directly the decrease in free energy resulting from the transfer of reactants into products at any arbitrarily chosen activities. For practical purposes, partial pressures, for gas reactions, or concentrations or mole fractions, for reactions in solution, may be used instead of activities, provided there are no serious deviations from ideal behavior. If the equilibrium constant of the reaction, or the standard free energy change ΔF^0 , which is equal to $-RT \ln K$, is known, the actual free energy change ΔF for the given conditions can then be calculated directly from the reaction isotherm (cf. p. 824). Some less obvious examples of this calculation are given below.

Suppose it is required to determine the change of free energy accompanying the combination of 1 mole of hydrogen and 0.5 mole of oxygen, each at 1 atm. pressure, to form liquid water at 25° c.; this is equivalent to the conversion of the oxygen and hydrogen at 1 atm. to water vapor at its saturation pressure at 25° c., i.e., 23.7 mm., or 23.7/760 atm., for the

water vapor will then be in equilibrium with the liquid, and no change of free energy is involved in the transfer from vapor to liquid. It follows, therefore, that the free energy change of the process

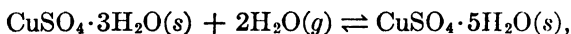


is given by

$$-\Delta F = RT \ln K_p - RT \ln \frac{23.7/760}{1 \times 1}, \quad (92)$$

where K_p is the equilibrium constant for the reaction between hydrogen, oxygen and water in the gaseous state, with partial pressures in atm.; for the present purpose the gases are assumed to behave ideally. The value of K_p cannot be determined directly from equilibrium measurements, since the degree of dissociation of water vapor is so very small at 25° c., but it can be estimated from the equation giving the variation of the observed equilibrium constant with temperature, assuming it to hold at low temperatures. In this manner K_p is found to be 1.01×10^{40} , a result which is probably fairly correct. Inserting this in (92), $-\Delta F$ may be evaluated as 56,570 cal. at 25°.* It appears, therefore, that the reaction between hydrogen and oxygen gases at 1 atm. pressure is accompanied by a considerable decrease of free energy at ordinary temperatures. The process should, therefore, be spontaneous, but it is so slow that it takes place at an appreciable rate only in the presence of a catalyst.

As an illustration of the calculation of the free energy change of a heterogeneous process, the free energy of hydration of a lower salt hydrate or of an anhydrous salt may be considered; in the system



the decrease of free energy resulting from the hydration of the trihydrate by 1 mole of liquid water at 25° is given by

$$-\Delta F = RT \ln 1/(p_{\text{H}_2\text{O}})_e - RT \ln 1/p_{\text{H}_2\text{O}}, \quad (93)$$

where $(p_{\text{H}_2\text{O}})_e$ is the equilibrium vapor pressure of the salt hydrate system and $p_{\text{H}_2\text{O}}$ is the saturation vapor pressure in equilibrium with liquid water at 25° c. The equilibrium constant (K_p) is here represented by $1/(p_{\text{H}_2\text{O}})_e$, since the water vapor appears on the left-hand side of the equation as written above. The value of $(p_{\text{H}_2\text{O}})_e$ is 7.8 mm., and that of $p_{\text{H}_2\text{O}}$ is 23.7 mm., and hence $-\Delta F$ is 657 cal. The free energy of hydration of various salts by ice at 0° c. has been evaluated in this manner by H. Schottky (1908), and the results compared with the heat of hydration ($-\Delta H$); as is to be expected, since $-\Delta F$ and $-\Delta H$ are not necessarily equal, the values do not coincide.

A somewhat different procedure is used to determine the free energy of transition of one modification of a substance, e.g., rhombic (α) sulfur, to another, e.g., monoclinic (β) sulfur, at a definite temperature. If the

* The best available value, derived indirectly, is 56,700 cal. at 25° c.

vapor pressures of the two forms are known at the given temperature, e.g., p_α and p_β , then assuming the vapor to obey the ideal gas laws, the decrease of free energy in transferring 1 mole from α - to β -sulfur, *via* the vapor, is

$$-\Delta F = \mu_\alpha - \mu_\beta = RT \ln p_\alpha/p_\beta. \quad (94)$$

Below the transition point (94° c.) the α -form is stable and hence has the lower vapor pressure; the quantity $-\Delta F$ has a negative value, i.e., ΔF is positive, and the spontaneous transfer from α (rhombic) to β (monoclinic) form is not possible. Above the transition point, however, p_α is greater than p_β , and there is a positive tendency for the transformation from α - to β -sulfur to occur, as indicated by the fact that ΔF is negative. At the transition point the two forms are in equilibrium; hence ΔF must be zero and both must have the same vapor pressure (cf. p. 469). If the substances involved in the transformation are not appreciably volatile, a method depending on the measurement of solubility may be employed. If s_α is the saturation solubility of one form in a given solvent and s_β that of the other form in the same solvent, at the same temperature, then

$$-\Delta F = \mu_\alpha - \mu_\beta = RT \ln s_\alpha/s_\beta \quad (95)$$

per mole of dissolved substance, provided the solutions are ideal. The solubilities should be expressed in mole fractions, but if the solutions are dilute, molar concentrations may be used without causing much error. Provided the solvent does not react in any way with the dissolved substance and the solutions are ideal, the value of $-\Delta F$ obtained from (95), and hence the ratio of the solubilities, should be independent of the nature of the solvent, at a given temperature.

Electrical Method for Determination of Free Energy Change.—The electrical method is one of the most accurate for the determination of free energy changes of reactions involving electrolytes, provided the process can be made to take place in a (thermodynamically) reversible cell (cf. p. 934). If E is the E.M.F. of the cell, and nF is the number of coulombs of electricity which must pass for the reaction to occur to the extent represented by the chemical equation, the electrical work is given by nFE . Since this is the whole of the reversible work, apart from that due to a volume change, it may be identified with the free energy decrease accompanying the reaction taking place in the cell, provided the E.M.F. is measured at constant e.g., atmospheric, pressure (cf. p. 230); hence, it is possible to write

$$-\Delta F = nFE. \quad (96)$$

Some reference to the use of this relationship is given on p. 868, but most of its applications will be considered in Chapter XII.

Standard Free Energies of Reactions.—As already seen, the standard free energy change of a reaction is important, because it is related to the equilibrium constant; from either of these the actual free energy change for a process, in which the reactants and products are not necessarily in

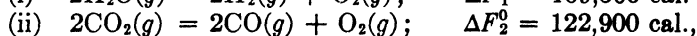
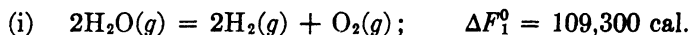
their standard states, can be evaluated. Like the equilibrium constant, the standard free energy change can conveniently be expressed as a function of the temperature, e.g., equation (65), as shown earlier (cf. p. 831). The use of standard free energies is particularly convenient for a number of reasons. For example, if ΔF^0 for a reaction is negative, it shows at once that the reaction is possible when the reactants and products are all in their standard states of unit activity. Further, standard free energy changes can be added and subtracted in the same manner as the change of heat content, i.e., the heat of reaction (p. 202), as may be readily shown. Consider the three reactions mentioned on p. 841; suppose the standard free energy changes are $-\Delta F_1^0$, $-\Delta F_2^0$ and $-\Delta F_3^0$, for the reactions (i), (ii) and (iii), respectively. The corresponding equilibrium constants are K_1 , K_2 and K_3 , where K_3 is equal to $\sqrt{K_2/K_1}$, so that

$$\ln K_3 = \frac{1}{2} \ln K_2 - \frac{1}{2} \ln K_1; \quad (97)$$

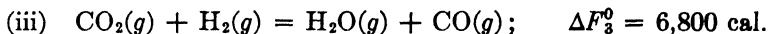
$$\therefore RT \ln K_3 = \frac{1}{2} RT \ln K_2 - \frac{1}{2} RT \ln K_1; \quad (98)$$

$$\therefore -\Delta F_3^0 = -\frac{1}{2}(\Delta F_2^0 - \Delta F_1^0). \quad (99)$$

If the equations are written out,* thus at 25° c.,

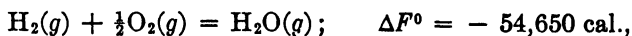


it is seen that if (i) and (ii) are each divided by two, and the former subtracted from the latter, then



This result, obtained by treating free energy equations as additive, like thermochemical equations, is exactly the same as the value for ΔF_3^0 derived from the theoretical equation (99). It is thus possible, by combining reactions for which the standard free energies are known, to calculate the values at the same temperature for other processes unsuitable for direct experimental study. The method is applicable to heterogeneous, as well as to homogeneous, reactions.

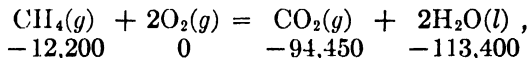
Free Energies of Formation.—Since absolute free energies are not known, but only differences between one state and another, it is desirable to adopt a convention, similar to that used for changes of heat content, in order to express standard free energies of formation of compounds (p. 202). The free energy of all elements in their standard states, in the form in which they exist at ordinary temperatures, is taken as zero. The standard free energy change of the reaction between hydrogen and oxygen at 25° c. may be represented as



* According to the usual convention, the free energy change is written as $+\Delta F^0$, just as the heat of reaction is expressed as $+\Delta H$ (p. 200).

the hydrogen, oxygen and water vapor being assumed to be in the ideal state at 1 atm. pressure. By convention ΔF^0 for hydrogen and oxygen gases are zero, so that the standard free energy of formation of water vapor is $-54,650$ cal. per mole. For liquid water at 25°C . the value of ΔF^0 is $-56,700$ cal. (cf. p. 854).

By the use of the methods described above, and others to be mentioned shortly, ΔF^0 data for a number of compounds at 25°C . have been tabulated, and with them, bearing in mind their additivity, the standard free energy changes for various reactions can be determined. Many of these are apparently not chemically reversible, that is to say, the reverse reaction occurs to such a minute extent as to be negligible; the free energy change could then not be evaluated directly, since the equilibrium constant cannot be determined. The method is exactly comparable with that employed in the calculation of the change of heat content (p. 202). Consider, for example, the reaction at 25°C .



the standard free energy of formation values being inserted below each substance; for the whole reaction, therefore,

$$\Delta F^0 = -113,400 + (-94,450) - (-12,200) = -195,650 \text{ cal.},$$

which may be compared with the heat content change of $-212,700$ cal. From a knowledge of ΔF^0 it is obviously possible to determine the equilibrium constant of the corresponding reaction, if it should be required.

A number of standard molar free energies of formation of compounds at 25°C ., in addition to those given above, are quoted in Table 144.

TABLE 144. STANDARD FREE ENERGIES OF FORMATION AT 25°C .

Gases	ΔF^0 kcal.	Solids and Liquids	ΔF^0 kcal.
Carbon monoxide	-33.0	Benzene (l)	29.4
Nitric oxide	20.7	Ethyl alcohol (l)	-40.2
Nitrogen dioxide	12.3	Acetic acid (l)	-94.5
Ammonia	-3.9	Silver chloride	-26.2
Sulfur dioxide	-71.7	Lead chloride	-75.0
Hydrogen sulfide	-7.9	Mercuric oxide	-13.9
Hydrogen chloride	-22.7	Lead monoxide	-45.0

THE THIRD LAW OF THERMODYNAMICS

The Nernst Heat Theorem.—By means of the integrated form of the van't Hoff equation (p. 830), it is possible to determine the change in heat content of a reaction, provided the equilibrium constant is known at two or more temperatures. Since these constants are related to the free energy change, it is evident that ΔH may be determined provided ΔF is available for two or more temperatures. It is, therefore, a direct result of the first and second laws of thermodynamics, upon which the van't Hoff

equation is based, that heat changes may be determined from free energy data. The same conclusion is reached directly from the Gibbs-Helmholtz equation (p. 232), from which it can be seen that ΔH can be calculated provided ΔF and $\partial(\Delta F)/\partial T$ are known,* the latter quantity being obviously equivalent to a knowledge of ΔF at two or more temperatures. The problem that the two laws of thermodynamics leaves unsolved, however, is the reverse calculation, namely, the evaluation of free energy changes from heat content and other thermal data. The question was discussed by H. Le Chatelier (1888), T. W. Richards (1902), and others, but the advance which led to a solution was made by W. Nernst (1906). From an examination of the Gibbs-Helmholtz equation

$$\Delta F - \Delta H = T \left(\frac{\partial(\Delta F)}{\partial T} \right)_P, \quad (100)$$

it is at once evident that unless $\partial(\Delta F)/\partial T$ is infinite ΔF must be equal to

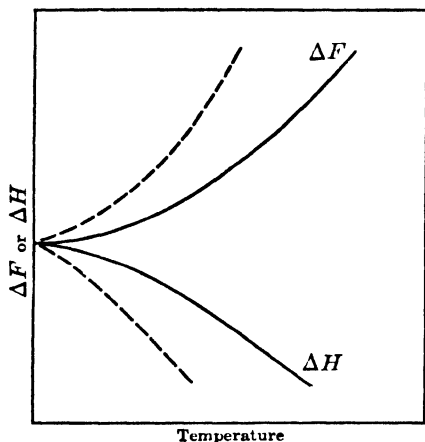


Fig. 226. The Nernst heat theorem

ΔH at the absolute zero, i.e., when $T = 0$. In the course of his observations on the E.M.F.'s of cells at low temperatures, Richards found that as the temperature is diminished the quantity $\partial(\Delta F)/\partial T$ decreases, so that ΔF and ΔH approach each other more closely, and the important suggestion was made by Nernst that the value of $\partial(\Delta F)/\partial T$ approaches zero asymptotically as the temperature is lowered toward the absolute zero. This means that ΔF and ΔH are not merely equal at 0°K. , but the values approach each other asymptotically in the vicinity of

that temperature. The mathematical expression of this postulate, generally known as the **Nernst heat theorem**, is

$$\lim_{T \rightarrow 0} \frac{d(\Delta F)}{dT} = \lim_{T \rightarrow 0} \frac{d(\Delta H)}{dT} = 0. \quad (101)$$

The curves representing the variation of ΔF and ΔH with temperature would be of the type shown by the full lines in Fig. 226, and not by the broken lines. In the diagram ΔF is depicted as greater than ΔH away from the absolute zero, but the reverse state of affairs is, of course, also possible since $\partial(\Delta F)/\partial T$ may be positive or negative. It has been shown

* It should be recalled that $(\partial(\Delta F)/\partial T)_P$ is equal to $-\Delta S$ (p. 232).

(pp. 232, 213) that

$$\left(\frac{\partial(\Delta F)}{\partial T}\right)_P = -\Delta S \quad \text{and} \quad \left(\frac{\partial(\Delta H)}{\partial T}\right)_P = \Delta C_P,$$

and so it is a consequence of the Nernst heat theorem that

$$\lim_{T \rightarrow 0} \Delta S = 0 \quad \text{and} \quad \lim_{T \rightarrow 0} \Delta C_P = 0. \quad (102)$$

Since gases do not exist at the absolute zero, the theorem is obviously not applicable to them, and there is reason for believing, contrary to Nernst's original views, that it does not hold for liquids; it must, therefore, be restricted to systems involving solids only.

The general integral of (101), p. 227, may be written in the form

$$\Delta S = \int_0^T \frac{\Delta C_P}{T} dT + I_0, \quad (103)$$

where I_0 is an integration constant. The quantity $\Delta C_P/T$ is known to be finite or zero, but according to the Nernst theorem ΔS approaches zero as the temperature is lowered, and this can only be true if I_0 is also zero. The equation for the entropy change at constant pressure, for a system of solids only, would thus become

$$\Delta S = \int_0^T \frac{\Delta C_P}{T} dT, \quad (104)$$

and since $\Delta F = \Delta H - T\Delta S$ (cf. p. 230), it follows that

$$\Delta F = \Delta H - T \int_0^T \frac{\Delta C_P}{T} dT, \quad (105)$$

thus providing a method for evaluating the free energy (ΔF) from a knowledge of the change of heat content (ΔH) and the heat capacity data included in the integrand.¹¹

In order to solve the integral, Nernst expressed the variation of ΔC_P with temperature in the usual manner, viz.,

$$\Delta C_P = \alpha + \beta T + \gamma T^2 + \dots, \quad (106)$$

so that

$$\int_0^T \frac{\Delta C_P}{T} dT = \alpha \ln T + \beta T + \frac{\gamma}{2} T^2 + \dots, \quad (107)$$

and therefore

$$\Delta F = \Delta H - \alpha T \ln T - \beta T^2 - \frac{\gamma}{2} T^3 - \dots, \quad (108)$$

It has been already seen (p. 214) that

$$\Delta H = \Delta H_0 + \alpha T + \frac{\beta}{2} T^2 + \frac{\gamma}{3} T^3 + \dots; \quad (109)$$

$$\therefore \Delta F = \Delta H_0 + \alpha T(1 - \ln T) - \frac{\beta}{2} T^2 - \frac{\gamma}{6} T^3 - \dots, \quad (110)$$

By the heat theorem $\partial(\Delta F)/\partial T$ must approach zero when $T = 0$, and differentiation of (110) with respect to temperature shows that α must consequently be zero; hence,

$$\Delta F = \Delta H_0 - \frac{\beta}{2} T^2 - \frac{\gamma}{6} T^3 - \dots \quad (111)$$

and

$$\Delta H = \Delta H_0 + \frac{\beta}{2} T^2 + \frac{\gamma}{3} T^3 + \dots \quad (112)$$

and

$$\Delta C_P = \beta T + \gamma T^2 + \dots \quad (113)$$

It appears that by means of these equations, free energy changes of systems involving solids can be calculated from thermal data. It must be pointed out, however, that in evaluating the integral it was tacitly assumed that (106) holds right down to the absolute zero; at low temperatures this is not true, since the heat capacity becomes proportional to T^3 (see p. 419), and consequently the treatment given above must be erroneous. This aspect of the development of the heat theorem should, therefore, be regarded as unsound; the agreement between experimental results and those obtained from (111) and (112), quoted by Nernst, is probably fortuitous. No doubt is, however, cast on the value of (105).

Application to Gaseous Systems.—The heat theorem applies to systems involving solids only, but Nernst showed that the conclusions could be extended to gas reactions. The variation of the vapor pressure of a pure solid p^0 with temperature can be represented by the Clapeyron-Clausius equation (p. 460); thus,

$$\frac{d \ln p^0}{dT} = \frac{L_s}{RT^2}, \quad (114)$$

where L_s is the molar heat of sublimation, the vapor being assumed ideal. The Kirchhoff equation (p. 213) may be written (approximately)

$$\frac{dL_s}{dT} = \Delta C'_P, \quad (115)$$

where $\Delta C'_P$ is equal to the heat capacity of the vapor minus that of the solid, at constant pressure; integration of (115) gives

$$L_s = L_{s(0)} + \int_0^T \Delta C'_P dT, \quad (116)$$

and on substituting in (114) and integrating, it follows that

$$\ln p^0 = -\frac{L_{s(0)}}{RT} + \int \frac{dT}{RT^2} \int_0^T \Delta C'_P dT + i, \quad (117)$$

the integration constant i being called the **true chemical constant** of the given substance. The second term on the right-hand side may be inte-

grated by parts; thus,

$$\int \frac{dT}{RT^2} \int_0^T \Delta C'_P dT = - \frac{\int_0^T \Delta C'_P dT}{RT} + \int_0^T \frac{\Delta C'_P}{RT} dT, \quad (118)$$

the first term on the right being, by (116), equal to $(L_s - L_{s(0)})/RT$, so that (117) becomes

$$\ln p^0 = - \frac{L_s}{RT} + \int_0^T \frac{\Delta C'_P}{RT} dT + i. \quad (119)$$

Consider the reaction $A(g) \rightleftharpoons B(g)$ occurring in the gas phase; a simple reaction has been chosen in order to avoid mathematical complexity. The equilibrium constant K_p is equal to p_B/p_A , where the p terms are the partial pressures at equilibrium; the van't Hoff equation is then

$$\frac{d \ln K_p}{dT} = \frac{\Delta H^G}{RT^2}, \quad (120)$$

where ΔH^G is the change of heat content in the gas reaction, and by Kirchhoff's equation

$$\frac{d(\Delta H^G)}{dT} = \Delta C_P^G = C_{P(B)}^G - C_{P(A)}^G, \quad (121)$$

the C_P^G terms being the heat capacities of the gases or vapors. By adopting a treatment identical with that just employed, it follows, by analogy with (119), that

$$\ln K_p = - \frac{\Delta H^G}{RT} + \int_0^T \frac{\Delta C_P^G}{RT} dT + i_R, \quad (122)$$

where i_R is the appropriate integration constant.

It is now necessary to determine the free energy increase accompanying the process $A(s) \rightarrow B(s)$ when both substances are solid; this may be evaluated with the aid of an equilibrium box, at the same temperature, in which the gases A and B react in the *absence* of solid phase. The vapor pressures of the solids are p_A^0 and p_B^0 , while the partial pressures in the equilibrium box are p_A and p_B , so that the free energy change ΔF^s for the reaction between the solids is

$$\Delta F^s = RT \ln \frac{p_A}{p_A^0} + RT \ln \frac{p_B^0}{p_B} \quad (123)$$

$$= RT \left(\ln \frac{p_B^0}{p_A^0} - \ln \frac{p_B}{p_A} \right) \quad (124)$$

$$= RT \left(\ln \frac{p_B^0}{p_A^0} - \ln K_p \right). \quad (125)$$

Assuming the vapors to behave as ideal gases, the expressions for the vapor pressures of the solids, i.e., p_A^0 and p_B^0 , are obtained from (119)

whereas that for K_p is given by (122), and if these are inserted in (125), it follows that

$$\Delta F^S = (L_{s(A)} - L_{s(B)} + \Delta H^G) + RT \int_0^T \frac{\Delta C'_{P(B)} - \Delta C'_{P(A)} - \Delta C_P^G}{RT} dT + RT(i_B - i_A - i_R). \quad (126)$$

The first term on the right-hand side, in the parentheses, is the increase of heat content in the reaction involving *solid* A and B, i.e., ΔH^S , and the numerator of the integrand is the difference in the heat capacities of A and B in the *solid* state, i.e., $-\Delta C_P^S$; hence (126) may be written

$$\Delta F^S = \Delta H^S - T \int_0^T \frac{\Delta C_P^S}{T} dT + RT(i_B - i_A - i_R). \quad (127)$$

If this equation is now compared with (105), which is applicable to a reaction involving solids only, it is immediately evident that

$$i_B - i_A - i_R = 0; \quad \therefore i_R = i_B - i_A, \quad (128)$$

so that the integration constant of (122) is equal to the algebraic sum ($\Sigma \nu i$) of the true chemical constants of the reacting substances; hence the Nernst heat theorem leads to the result

$$\ln K_p = -\frac{\Delta H^G}{RT} + \int_0^T \frac{\Delta C_P^G}{RT} dT + \Sigma \nu i, \quad (129)$$

where K_p is the equilibrium constant for the *gas* reaction. The true chemical constants may, theoretically, be determined from vapor pressure and heat capacity measurements, by means of (119), and so it becomes possible to evaluate K_p for a gas reaction, and consequently the free energy change, from thermal and vapor pressure data. Although (129) has been developed for a very simple reaction, exactly the same result may be obtained, by following the same treatment, for a more complex process. In the deduction of this equation a homogeneous gas reaction was assumed; if the process is heterogeneous, however, then the quantity $\Sigma \nu i$ involves terms only for those reactants and products present as gases, and not for the solid reacting species. This may be proved by considering the reaction $A(g) + S(s) \rightleftharpoons B(g)$, where S is present as solid; the form of the equation for K_p remains unchanged, and no additional term appears in (126), so that, as before, i_R is found to be equal to $i_B - i_A$.

The Nernst Approximation Formula.—The application of (129) is obviously difficult, since it would be necessary to have data of the heat capacities and vapor pressures down to the absolute zero; an approximate form has, therefore, been derived by Nernst in the following manner. From the Clapeyron-Clausius equation (p. 460) in the form

$$\frac{dp^0}{dT} = \frac{L_s}{T(v^G - v^S)} \quad (130)$$

for the sublimation of a pure solid, and the empirical equation

$$p^0(v^G - v^S) = RT(1 - p^0/p_c), \quad (131)$$

where p^0 is the vapor pressure of the solid and p_c its critical pressure, there can be obtained

$$\frac{d \ln p^0}{dT} \left(1 - \frac{p^0}{p_c} \right) = \frac{L_s}{RT^2}. \quad (132)$$

Replacing L_s , the heat of sublimation, by the empirical expression

$$L_s = (L_{s(0)} + aT - bT^2)(1 - p^0/p_c) \quad (133)$$

and integrating, it follows that

$$\ln p^0 = -\frac{L_{s(0)}}{RT} + \frac{a}{R} \ln T - \frac{b}{R} T + i, \quad (134)$$

where i should be, at least approximately, equal to the true chemical constant; (134) is, in fact, seen to be an approximate form of (119). From measurements of the heat capacities of gases Nernst concluded that the value 3.5 may be used for the constant a for all substances, so that a/R is approximately 1.75, and converting natural to common logarithms, (134) becomes

$$\log p^0 = -\frac{L_{s(0)}}{4.57T} + 1.75 \log T - \frac{b}{4.57} T + I. \quad (135)$$

The constant I might be, at first sight, supposed to equal $i/2.303$, but as a consequence of the various approximations introduced it is virtually a new empirical constant, and so it is called the **conventional chemical constant**. Values have been determined directly, in certain instances, from vapor pressure measurements by means of (135), and in addition the constants for a number of gases have been estimated from empirical relationships discovered to exist between I and other physical properties, e.g., boiling point, critical pressure, and vapor pressure near the critical point. Some of the results obtained by Nernst, based on pressures expressed in atm., are recorded in Table 145; except for hydrogen (and helium) they seem to fall between 2.5 and 4.

TABLE 145. CONVENTIONAL CHEMICAL CONSTANTS

Hydrogen	1.6	Chlorine	3.1	Nitric oxide	3.5
Methane	2.5	Carbon dioxide	3.2	Carbon monoxide	3.5
Nitrogen	2.6	Ammonia	3.3	Bromine	3.5
Oxygen	2.8	Hydrogen bromide	3.3	Water	3.6
Hydrogen chloride	3.0	Hydrogen iodide	3.4	Iodine	3.9

By a similar method to that used above, it is possible to deduce an equation for K_p for a gas reaction analogous to that for the vapor pressure [cf. equations (119) and (122)]; hence,

$$\log K_p = -\frac{\Delta H_0^G}{4.57T} + \Sigma \nu 1.75 \log T + \frac{\beta}{4.57} T + \Sigma \nu I, \quad (136)$$

where the variation of ΔH^G with temperature is given by

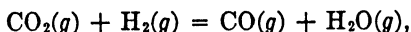
$$\Delta H^G = \Delta H_0^G + \Sigma \nu 3.5T + \beta T^2 \quad (137)$$

and $\Sigma \nu$ refers to *gaseous* reactants only. The equation (136) is admittedly approximate, and Nernst proposed to simplify it still further by neglecting the term involving β , and replacing ΔH_0^G by ΔH^G , the value at ordinary temperatures,

thus giving the **Nernst approximation formula** for gas reactions

$$\log K_p = -\frac{\Delta H}{4.57T} + \Sigma \nu 1.75 \log T + \Sigma \nu I. \quad (138)$$

In spite of its very approximate nature, (138) appears to give reasonably good results for certain cases, particularly when $\Sigma \nu$ is zero, probably because of the partial cancellation of errors. Consider, for example, the water gas reaction



for which $\Sigma \nu I$ is $(3.5 + 3.6) - (3.2 + 1.6) = 2.3$, and ΔH is 10,000 cal.; hence,

$$\log K_p = -\frac{10,000}{4.57T} + 2.3. \quad (139)$$

At 1073° K., $\log K_p$ is thus 0.26 and so K_p is 1.82, in fair agreement with the value given in Table 142, p. 851. Similar agreement has been found for other reactions, e.g., dissociation of hydrogen chloride, bromide and iodide, and of nitric oxide; in all these cases $\Sigma \nu$ is zero, that is, there is no change in the number of molecules as a result of the reaction.

The Nernst formula also gives fairly good values for the dissociation pressures of carbonates and oxides; these are both heterogeneous processes involving one gaseous species, so that $\Sigma \nu$ is 1. The respective equations are

$$\log p_{\text{CO}_2} = -\frac{\Delta H}{4.57T} + 1.75 \log T + 3.2 \quad (140)$$

and

$$\log p_{\text{O}_2} = -\frac{\Delta H}{4.57T} + 1.75 \log T + 2.8. \quad (141)$$

By means of these equations it is possible to calculate the temperature at which the dissociation pressures of various carbonates, oxides, etc., become equal to 1 atm., provided the heats of dissociation are known; the results have been found to be in agreement with those obtained experimentally. It will be noted, incidentally, that (140) is in agreement with the general rule (p. 848) that the temperature at which the dissociation pressure of a carbonate is equal to 1 atm. increases as the heat of dissociation increases.

Third Law of Thermodynamics.—The suggestions on which the Nernst heat theorem was based have been developed so as to lead to a new law of experience that has been called the **third law of thermodynamics**. It has been seen that by the heat theorem ΔC_P approaches zero at 0° K., so that at the absolute zero the heat capacities of reactants and products are identical. It appears highly probable, therefore, that at 0° K. all substances have the same heat capacity. The application of the quantum theory to the interpretation of the specific heats of solids (p. 417) indicated that heat capacities C_P of solids should tend towards zero at the absolute zero of temperature, and that C_P/T should be either zero or finite. These conclusions have been borne out by experiment, and so it is possible to extend the heat theorem by writing

$$\lim_{T \rightarrow 0} C_P = 0. \quad (142)$$

Further, the Nernst theorem states that in the limit ΔS becomes zero for solids, so that in the solid state all substances should have the same entropy at the absolute zero. The difference in entropy between a substance at 0°K. and at $T^\circ \text{K.}$ is given by (cf. p. 227)

$$S_T - S_0 = \int_0^T \frac{C_P}{T} dT, \quad (143)$$

and since C_P/T is finite or zero, it follows that $S_T - S_0$ is always positive or zero; in other words, a substance has its lowest possible entropy value at the absolute zero of temperature. In conjunction with the heat theorem, this fact led M. Planck (1912) to suggest that for *pure* solids and liquids the entropy approaches zero at 0°K. ; this generalization appears to be true for solids, but not necessarily for liquids, and it has led to the following statement of the third law of thermodynamics. *Every substance has a finite positive entropy, but at the absolute zero of temperature the entropy may become zero, and does so become in the case of a perfect crystalline substance.*

From experimental evidence it is probable that, except for liquid helium, the law does not apply to (supercooled) liquids; neither does it apply to solid solutions. If x_A and x_B are the mole fractions of two constituents A and B in such a solution, assumed to be ideal, then the change of free energy resulting from the transfer of x_A moles of A and x_B moles of B from the pure solid to the solid solution is given by

$$\Delta F = x_A RT \ln x_A + x_B RT \ln x_B. \quad (144)$$

It has been seen (p. 232) that $-\Delta S = (\partial(\Delta F)/\partial T)_P$; hence on differentiation it follows that

$$-\Delta S = x_A R \ln x_A + x_B R \ln x_B, \quad (145)$$

which gives the difference of entropy between the pure solids and the solid solution.* Since x_A and x_B are fractional the logarithms are negative, and so ΔS is positive; the solid solution therefore has the higher entropy and hence cannot be zero at 0°K. The entropy of a system in any state may be regarded as a measure of the "randomness" of that state (p. 229). In the perfect crystal of a pure substance at the absolute zero there is complete order, and hence the entropy is zero. In a liquid or a solid solution there must be some randomness in the arrangement of the atoms or molecules, and so the entropy at 0°K. is greater than for the pure solid.

The third law includes the heat theorem, but it has a great advantage over the latter in the matter of its exact application. It has been seen that Nernst was forced to use approximation methods, chiefly because of the difficulty of the direct determination of the true chemical constants from vapor pressure data at low temperatures. By means of the third law, however, the constant can be obtained indirectly from heat capacity data, although it has become the practice with modern workers not to

* Equation (145) is the form taken by (94), p. 226, for the molar entropy of mixing for a homogeneous system of two components.

evaluate the chemical constants, but to express observations in terms of the entropies of various substances at 25° c., i.e., 298.16° K. With the aid of the entropy and heat content data it is possible to calculate the free energy changes and equilibrium constants for reactions at this temperature; further, if the heat capacities of the reacting substances are known over a range of temperature, the results may be extended to any temperature within this range. The effect of the third law of thermodynamics has thus been to concentrate interest on the determination of absolute entropies.¹²

Determination of Entropies by Third Law.—If the entropy of a pure solid at 0° K. is zero, then at any other temperature the entropy is given by

$$S_T = \int_0^T \frac{C_P}{T} dT = \int_0^T C_P d \ln T, \quad (146)$$

and hence the evaluation of S_T requires accurate specific heat data at low temperatures. Special experimental methods, involving the use of vacuum-jacketed calorimeters, have been developed in order to reduce heat losses by radiation. An accurately known amount of heat is produced by means of an electric current and the consequent rise of temperature is measured with a very sensitive resistance thermometer or thermocouple. Since measurement cannot be made right down to 0° K., the observations are carried to as low a temperature as possible, and from these the function θ of the Debye specific heat equation (p. 419) is estimated; the values of C_V at low temperatures can thus be calculated and these must be converted to C_P before evaluating the entropy by (146).*

The integration of (146) is usually carried out graphically, from the plot of C_P/T against T , or of C_P against $\ln T$, as in Fig. 227. The heat capacities plotted on the curve BC are the experimental values, the point B corresponding to the lowest temperature, e.g., 10° to 15° K., at which satisfactory measurements are possible. The portion AB , down to the absolute zero, is derived from the Debye equation; the heat capacity of a solid approaches zero rapidly at temperatures below 10° K., and so this part of the curve has little influence on the accuracy of the results. The area enclosed by the curve ABC , the $\ln T$ axis and the ordinate for any given temperature gives the entropy of the solid at that temperature. If C is the melting point of the solid, there is a sharp change in the heat capacity, and DE is the plot of C_P against $\ln T$ for the liquid. The change of state at the melting point is accompanied by absorption of the latent heat of fusion and there is a corresponding entropy increase given by L_f/T (p. 462), where L_f is the heat absorbed per mole and T is the

* Equation (146) may be written

$$S_T = \int_0^T C_P d \ln T = \int_0^T C_V d \ln T + \int_0^T (C_P - C_V) d \ln T.$$

and since $C_P - C_V \approx aT^{3/2}$ (see p. 417), the last integral is equal to $\frac{2}{5}(C_P - C_V)$. At low temperatures $C_P - C_V$ is small, and so the difference between the integrals of $C_P d \ln T$ and $C_V d \ln T$ may be neglected.

melting point. The entropy increase resulting from raising the temperature of the liquid to its boiling point at 1 atm. pressure is obtained from the area under the curve in Fig. 227 between melting point and boiling

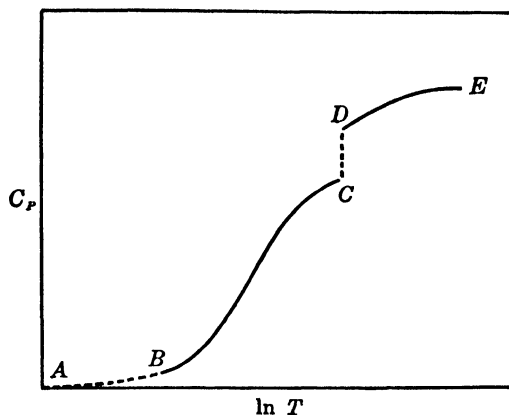


FIG. 227. Graphical determination of entropy change

point. The entropy of vaporization is then added to give the total for the gas at its boiling point under 1 atm. pressure. If the solid undergoes a polymorphic change, as is frequently the case, the appropriate allowance must be made for the entropy increase; it is equal, as for other changes of state, to the heat of transition per mole divided by the transition temperature. In order to illustrate the above description the results obtained by W. F. Giaque and J. O. Clayton (1933) for nitrogen are given in Table 146. The Debye function was found, from low tempera-

TABLE 146. THE ENTROPY OF NITROGEN GAS

	Cal./degree
0° to 10° K. from Debye equation	0.458
10° to 35.61° K. (transition point) by graphical integration	6.034
Transition, 54.71 cal./35.61°	1.536
35.61° to 63.14° K. (melting point) by graphical integration	5.589
Fusion, 172.3 cal./63.14°	2.729
63.14° to 77.32° K. (boiling point) by graphical integration	2.728
Vaporization, 1332.9 cal./77.32°	17.239
Total	36.31

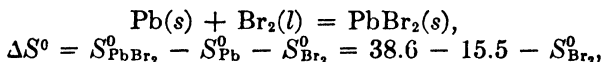
ture measurements, to be 68, and this was used to obtain the entropy at 10° K. At 35.61° K. the solid nitrogen undergoes a change of form, and the heat absorbed is 54.71 cal. per mole; the melting point is 63.14° K. and the molar heat of fusion is 172.3 cal.; and the boiling point at 1 atm. pressure is 77.32° K. with a molar heat of vaporization of 1332.9 cal. The entropy of gaseous nitrogen at its boiling point is thus 36.31 cal. per de-

gree, that is, entropy units, abbreviated to e.u. (p. 224), per mole. For purposes of calculating free energies it is desirable to have the entropy in the standard state, that is, for the ideal gas at 1 atm.; it is necessary to know, therefore, the change of entropy accompanying expansion of the *actual* gas from 1 atm. to zero pressure and the compression of an *ideal* gas back to 1 atm. In order to do this use may be made of the thermodynamic equation (p. 236)

$$\left(\frac{\partial S}{\partial P}\right)_T = -\left(\frac{\partial V}{\partial T}\right)_P, \quad (147)$$

together with a suitable equation of state for the actual gas; that of Berthelot (p. 295) is particularly convenient for this purpose, and the entropy of an actual gas is found to be $27RT^3P/32T^3P_c$ lower than that of an ideal gas, where P_c and T_c are the critical constants of the gas. For nitrogen at its boiling point the correction is 0.22 e.u. per mole giving 36.53 e.u. for the ideal gas. The entropy at 298.16° K. can be obtained by further addition of the integral of $(C_P/T)dT$ for the gas between 77.32° K. and 298.16° K. Highly accurate experimental data are, unfortunately, not available but a satisfactory estimate, viz., 9.36 e.u. can be made by statistical methods (see p. 870), so that the entropy of nitrogen as an ideal gas at 1 atm. pressure is $36.53 + 9.36 = 45.89$ e.u. per mole at 298.16° K.

The standard entropies of a number of elements and compounds have been ascertained by this method, which is the most reliable experimental procedure, based on the third law of thermodynamics, but others have also been used. If the entropies of a binary compound and of one of its constituent elements are known, that of the other can be evaluated if certain free energy and heat content data are available. The entropies of lead bromide and of lead, for example, obtained by direct calorimetric measurements, are 38.6 e.u. per mole and 15.5 e.u. per g. atom, respectively, in their standard states * at 298° K. For the reaction



so that if ΔS^0 for the reaction were known the entropy of the liquid bromine, $S_{\text{Br}_2}^0$, could be obtained. From measurements of the e.m.f.'s of cells (cf. p. 855) the free energy change ΔF^0 of the reaction at 298° K. is known to be -62,000 cal., and from thermal data the change of heat content † (ΔH) accompanying the process is approximately -66,000 cal. For a reaction in which all the substances are in their respective standard states, (113a), p. 230, becomes $\Delta F^0 - \Delta H^0 = -T\Delta S^0$, and since ΔH^0 may

* The standard state of a solid is taken as the pure solid and that of a liquid is the pure liquid.

† The value ΔH^0 for the standard state should be used, but the difference, which is in any case very small, is uncertain and may be neglected in comparison with errors in ΔH .

be replaced by ΔH without serious error,

$$\Delta F^0 - \Delta H = -T\Delta S^0. \quad (148)$$

From this expression ΔS^0 is found to be $-(66,000 - 62,000)/298^\circ$, i.e., -13.42 E.U. at 298°K. , so that $S_{\text{Br}_2}^0$ is 36.5 E.U. per mole. This result may be compared with the value 36.8 E.U. per mole obtained by the direct calorimetric method. For many reactions ΔF^0 can be derived from measurement of the equilibrium constant, and the method described above may then be applied to evaluate the standard entropy of one of the reactants if the values for the others are known.

The standard entropies of some solids and liquids at 25°C. are quoted in Table 147; these are all based on the assumption of the third law.*

TABLE 147. ENTROPIES OF SOLIDS AND LIQUIDS IN THEIR STANDARD STATES AT 25°C.

Elements				Compounds			
C (diamond)	0.59	Fe	6.5	NaF	12.4	AgCl	23.0
C (graphite)	1.36	Zn	9.9	NaCl	17.2	AgI	27.6
Na	12.25	Br ₂ (l)	36.8	KF	15.8	Hg ₂ Cl ₂	47.0
Mg	7.76	Ag	10.2	KCl	19.9	PbBr ₂	38.6
Al	6.73	I ₂ (s)	27.9	KBr	22.5	MgO	6.6
K	15.2	Hg(l)	18.5	KI	23.4	Al ₂ O ₃	12.5
Cu	8.0	Pb	15.5	H ₂ O(l)	16.7	HgO	10.5

Data for gases are recorded on p. 878. The standard entropies of many organic compounds have been determined, and utilized for the calculation of the free energies of reactions for which heat data are available (see p. 879).

Evidence for the Third Law.—The agreement between the two values for the entropy of bromine recorded above can be regarded as evidence of the reliability of the third law of thermodynamics. Alternatively, the third law entropies may be used in conjunction with thermochemical data to calculate the free energy of the formation of lead bromide, which may then be compared with the experimental value. Thus, taking the data based on the assumption that the entropy of a solid is zero at 0°K. , then at 298°K. , the entropies of 1 g. atom of lead, 1 mole of liquid bromine and 1 mole of lead bromide are 15.5, 36.8 and 38.6 E.U., respectively, so that ΔS^0 for the formation of lead bromide is $38.6 - 15.5 - 36.8 = -13.7$ E.U. per mole. The quantity $\Delta F^0 - \Delta H^0$ is, therefore, $13.7 \times 298 = 4080$ cal., and if ΔH^0 is $-66,000$ cal. then ΔF^0 should be $-61,920$ cal., in excellent agreement with $-62,000$ cal. obtained by direct experiment.

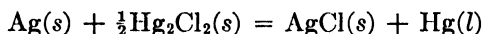
One of the chief difficulties involved in this method of verification lies in the uncertainty of the ΔH values in the literature; the following procedure has been adopted to overcome this source of error. According to

* Except for $\text{H}_2\text{O}(l)$ which has been corrected for the entropy of ice at 0°K. (see p. 877).

the Gibbs-Helmholtz equation, $\Delta F - \Delta H$ is equal to $-T(\partial(\Delta F)/\partial T)_P$, and since ΔF for a given reaction is equal to $-nFE$, where E is the E.M.F. of a reversible cell in which the reaction occurs, it follows that $\Delta F - \Delta H$ is equal to $nFT(\partial E/\partial T)_P$. Utilizing (113a), p. 230, it is seen, therefore, that

$$-T\Delta S = nFT(\partial E/\partial T)_P = \Delta F - \Delta H. \quad (149)$$

If $\Delta F - \Delta H$, determined from $T\Delta S$ on the basis of the third law, is in agreement with $nFT(\partial E/\partial T)_P$ obtained from E.M.F. measurements, the law may be regarded as substantiated in these instances. For the reaction



the value of $\Delta F - \Delta H$ from E.M.F. data is -2320 cal.* at 298°K . (R. H. Gerke, 1922); the standard entropies of the reacting substances at 298°K . are

$$S_{\text{Ag}}^0 = 10.2; \quad S_{\frac{1}{2}\text{Hg}_2\text{Cl}_2}^0 = 23.5; \quad S_{\text{AgCl}}^0 = 23.0; \quad S_{\text{Hg}}^0 = 18.5 \text{ E.U.};$$

$$\therefore \Delta S_{298}^0 = 7.8 \text{ E.U.}, \quad \text{and} \quad -T\Delta S^0 = -2324 \text{ cal.}$$

The agreement in this and similar cases provides evidence in favor of the third law of thermodynamics.¹³

It will be apparent that if the increase in heat content for any reaction is known, the standard free energy change can be derived from the entropies of the substances concerned. An illustration, for a process involving gases, is given on p. 878, and further applications to the calculation of electrode potentials and E.M.F.'s, which as already noted are related to the free energy of the reaction occurring in the cell, will be considered in Chapter XII.

Entropy by Statistical Methods.—The development of the statistical method (see p. 874) for the determination of entropies has not only served to confirm the third law of thermodynamics in many instances, but has also brought to light some important limitations of the law in its simple form. The heat capacity of a gas at constant volume can be expressed in terms of the partition function (p. 333) by means of (276) on page 333, viz.,

$$C_V = \frac{\partial}{\partial T} \left[RT^2 \left(\frac{d \ln Q}{dT} \right) \right]_V,$$

where Q is the partition function. Further, integration of (98) on page 226 gives

$$S - S_0 = \int_0^T \frac{C_V}{T} dT$$

for the difference between the entropy S of a substance at temperature T

* Since the reaction involves pure solids and liquids only, at 1 atm. pressure, this result refers to the standard states.

and the value S_0 at the absolute zero, at constant volume.* By combining these two equations, the result is

$$\begin{aligned} S - S_0 &= \int_0^T \frac{1}{T} \cdot \frac{\partial}{\partial T} \left[RT^2 \left(\frac{d \ln Q}{dT} \right) \right]_v dT \\ &= RT \left(\frac{\partial \ln Q}{\partial T} \right)_v + R \ln Q - R \ln Q_0, \end{aligned} \quad (150)$$

where Q_0 is the hypothetical partition function of one molecule of the gas at the absolute zero.

In order to eliminate S_0 use is made of the Planck modification of Boltzmann's postulate [equation (111), p. 229] relating the entropy to the thermodynamic probability W , i.e., $S = k \ln W$, where k is the gas constant per single molecule (p. 264). If a gas could exist at the absolute zero, all the molecules would be in the state of lowest energy, that is, in the zero-point energy level. The thermodynamic probability for a single molecule would then be given by the multiplicity or statistical weight (p. 95) of this energy state, i.e., p_0 . For N distinguishable molecules the total probability would be p_0^N , but since the molecules are not distinguishable, as explained on page 273, this probability must be divided by $N!$, so that the entropy of the N molecules of gas, i.e., S_0 , at the absolute zero would be given by

$$S_0 = k \ln \frac{p_0^N}{N!}. \quad (151)$$

If N is the Avogadro number, this gives the molar entropy, so that kN may be replaced by the molar gas constant R ; (151) then becomes

$$S_0 = R \ln p_0 - k \ln N! \quad (152)$$

As stated above, all the molecules would be in the zero-point energy level at the absolute zero; the partition function for a single molecule, i.e., Q_0 , would then be merely $p_0 e^{-0/kT}$, i.e., p_0 . Substitution in (152) then gives

$$S_0 = R \ln Q_0 - k \ln N!$$

and combination with (150) leads to the result

$$S = RT \left(\frac{\partial \ln Q}{\partial T} \right)_v + R \ln Q - k \ln N! \quad (153)$$

This equation is of great importance for it permits the evaluation of the *absolute entropy* of a mole of gas from a knowledge of the partition function. The result obtained in this manner is independent of the third law of thermodynamics, but it involves the postulate concerning entropy and probability.

* Constant volume is specified, since this condition is implicit in the subsequent treatment.

By making the assumption that the different forms of energy do not interact with one another (cf. p. 333), the complete partition function Q may be taken as the product of the partition functions $Q_{tr.}$ and $Q_{int.}$ for translational and internal, i.e., rotational and vibrational energies, respectively. The entropy may correspondingly be written as the sum of $S_{tr.}$ and $S_{int.}$, so that (153) may be divided into two parts, viz.,

$$S_{tr.} = RT \left(\frac{\partial \ln Q_{tr.}}{\partial T} \right)_v + R \ln Q_{tr.} - k \ln N! \quad (154)$$

$$S_{int.} = RT \frac{d \ln Q_{int.}}{dT} + R \ln Q_{int.} \quad (155)$$

It will be noted that the term $k \ln N!$ is included with the translational entropy; the presence of this term is due to the indistinguishability of the molecules, and this is only of significance when translational motion is considered. As far as rotation and vibration are concerned, the molecules may be treated as occupying fixed positions, and hence as being distinguishable. Further, the condition of constant volume has been omitted in the expression of the entropy corresponding to the internal energy, for the partition function is independent of the volume. The partition functions for rotation and vibration were considered in Chapter IV [cf. equations (281) and (283)], and that for translational energy will be developed below.

Translational Partition Function.—Imagine a single particle, e.g., a gas molecule, inside a rectangular box whose linear dimensions are a , b and c ; the potential energy inside the box may be taken as zero, that is, provided the x coordinate, for example, lies between 0 and a , the y coordinate between 0 and b , and the z coordinate between 0 and c . If the particle has a mass m , then application of the Schrödinger wave equation (p. 75), remembering that U is zero, gives

$$-\frac{\hbar^2}{8\pi m} \left(\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} \right) = \epsilon \psi, \quad (156)$$

where ψ is a function of the coordinates x , y and z ; ϵ is the translational energy of a single molecule. The assumption is now made that ψ can be represented by the product XYZ , where X , Y and Z are respectively functions of x , y and z only. Dividing through (156) by ψ , and substitution by XYZ then gives

$$-\frac{\hbar^2}{8\pi m} \left(\frac{1}{X} \cdot \frac{\partial^2 X}{\partial x^2} + \frac{1}{Y} \cdot \frac{\partial^2 Y}{\partial y^2} + \frac{1}{Z} \cdot \frac{\partial^2 Z}{\partial z^2} \right) = \epsilon. \quad (157)$$

It can be readily seen that since each of the three terms in the parentheses is a function of x , y or z only, and these coordinates are independent, the energy ϵ can be divided into three parts, ϵ_x , ϵ_y and ϵ_z , representing the amounts of translational energy of the particle parallel to the three axes. For the x coordinate, for example,

$$\epsilon_x = -\frac{\hbar^2}{8\pi m} \left(\frac{1}{X} \cdot \frac{\partial^2 X}{\partial x^2} \right), \quad (158)$$

and the general solution of this differential equation is

$$X = C \sin (Ax + B), \quad (159)$$

where A , B and C are constants. The probability of finding the particle at any point within the box is given by ψ^2 (see p. 76) and so X^2 is the probability of finding it along the x axis; this should be zero in the walls of the box, that is, when $x = 0$ and $x = a$, since these are the coordinates of the two walls perpendicular to the x axis. These conditions can only be satisfied if $A = n\pi/a$ and $B = 0$, so that (159) becomes, for this particular case,

$$X = C \sin \frac{n\pi}{a} x. \quad (160)$$

Substitution of this value of X in (156) gives for the energy in the direction parallel to the x axis

$$\epsilon_x = \frac{n^2 h^2}{8a^2 m}, \quad (161)$$

where n is an integer which may be regarded as the quantum number for translational energy in the x direction.

By the definition of the partition function (p. 333), the translational contribution may be written

$$Q_{tr.} = \sum_{n=0}^{\infty} e^{-n^2 h^2 / 8a^2 m k T}, \quad (162)$$

the statistical weight, i.e., p , of each level being unity. Since the translational energy levels are very closely spaced it is possible to replace summation by integration, so that

$$Q_{tr.} = \int_0^{\infty} e^{-n^2 h^2 / 8a^2 m k T} dn = \frac{(2\pi m k T)^{\frac{1}{2}}}{h} a. \quad (163)$$

The translational partition function, for motion along one axis, that is for one degree of freedom, is thus given by (163), and the corresponding value for all three degrees of translational movement of a single particle is the product of the three separate partition functions; thus,

$$Q_{tr.} = \frac{(2\pi m k T)^{\frac{3}{2}}}{h^3} V, \quad (164)$$

where V is the volume of the box inhabited by the particle. This is the expression generally employed for the translational contribution to the partition function of a single molecule of mass m , present in a volume V at a temperature T

Entropy of Monatomic Molecules: The Sackur-Tetrode Equation.—Apart from the possible influence of higher electronic states, which may be ignored at ordinary temperatures, the complete partition function of a monatomic gas is identical with that due to translation only. Insertion of (164) in (154) will therefore give the total molar entropy of a monatomic gas or the translational contribution to the entropy of any

ideal gas; thus,

$$S_{tr.} = \frac{3}{2}R + R \ln \frac{(2\pi mkT)^{\frac{3}{2}}}{h^3} V - k \ln N! \quad (165)$$

Since N is a large number, viz., the Avogadro number, it is possible to use the Stirling approximation and to replace $\ln N!$ by $N \ln N - N$; (165) then becomes

$$S_{tr.} = R \left[\ln \frac{(2\pi mkT)^{\frac{3}{2}}}{h^3 N} V + \frac{5}{2} \right]. \quad (166)$$

This equation, deduced in various forms by O. Sackur (1911-13), O. Stern (1913), H. Tetrode (1914), and others, is known as the **Sackur-Tetrode equation**. If the values of N , k , and V (in cc.) for an ideal gas, and the accepted figures for π and h are inserted in (166), it is found that

$$S_{tr.}^0 = \frac{3}{2}R \ln M + \frac{5}{2}R \ln T - 2.311, \quad (167)$$

where $S_{tr.}^0$ represents the entropy in the standard state, i.e., for 1 mole of the ideal gas at 1 atm. press.; M is now the ordinary molecular weight. If R is taken as 1.987 cal. deg.⁻¹ mole⁻¹, (167) gives the standard translational entropy in e.u. per mole. The entropies of helium, argon, mercury and cadmium vapors, all of which are monatomic, calculated in this manner are almost identical with the values obtained from specific heat data, assuming the third law to hold.¹⁴

Internal Entropy of Polyatomic Molecules.—For a polyatomic molecule the complete entropy is obtained by adding to $S_{tr.}^0$ the internal entropy $S_{int.}^0$ corresponding to energy of all types other than translational, i.e., electronic, vibrational and rotational. The internal entropy may be evaluated by means of (155) where the partition function Q is equal to $\sum p e^{-\epsilon/kT}$ (see p. 333) for the internal forms of energy. Insertion of this quantity in (155) gives

$$S_{int.}^0 = R \left(\ln \sum p e^{-\epsilon/kT} + \frac{1}{kT} \cdot \frac{\sum \epsilon p e^{-\epsilon/kT}}{\sum p e^{-\epsilon/kT}} \right), \quad (168)^*$$

from which $S_{int.}^0$ may be evaluated if the necessary rotational and vibrational energy values are known from spectroscopic measurements (cf. p. 563). The computations are very laborious, however, and an approximate method may be used which gives satisfactory results at ordinary temperatures. At such temperatures electronic and vibrational energies can be neglected since very few molecules will be in any but the lowest energy levels, and it is possible to write $Q_{int.}^0 \approx Q_{rot.}^0$, so that the rotational levels only may be assumed to contribute to the partition function.

It has been seen (p. 336) that, as a first approximation, the value of $Q_{rot.}$ for a *diatomic molecule* can be expressed as $8\pi^2 I kT/h^2$ multiplied by a

* The molar (internal) entropies given by (168) apply to the standard state, because they are independent of the volume (or pressure).

nuclear spin factor; thus,

$$(i) \text{ Symmetrical molecule: } Q_{\text{rot.}} = \frac{1}{2}(2i + 1)^2 8\pi^2 I k T / h^2, \quad (169)$$

$$(ii) \text{ Unsymmetrical molecule: } Q_{\text{rot.}} = (2i + 1)(2i' + 1) 8\pi^2 I k T / h^2, \quad (170)$$

where I is the moment of inertia. Since (169) and (170) each involve T , it follows that $d \ln Q_{\text{rot.}} / dT = 1/T$, and hence (155) now becomes

$$S_{\text{rot.}}^0 = R(\ln Q_{\text{rot.}} + 1). \quad (171)$$

Inserting the usual values for the universal constants π , k and h in (169) and (170) and substituting in (171), the corresponding results are

$$(i) \quad S_{\text{rot.}}^0 = R \ln IT + 177.7 + R \ln \frac{1}{2}(2i + 1)^2 \quad (172)$$

and

$$(ii) \quad S_{\text{rot.}}^0 = R \ln IT + 177.7 + R \ln (2i + 1)(2i' + 1). \quad (173)$$

These equations give the rotational entropy including the nuclear spin contribution. It is the general custom, however, to omit the latter because, with the exception of hydrogen and deuterium molecules, it is not included in the entropies obtained from thermal data. The reason is that the effect of the nuclear spin influences the heat capacity only at temperatures below those at which measurements can be made, and so in extrapolating from the observed values down to 0°K . the nuclear spin contribution is completely ignored. It can be shown that for practical purposes, if entropies are to be used for the calculation of free energy changes, the omission of the nuclear spin effect is immaterial, provided it is deliberately omitted in all cases, including those of hydrogen and deuterium where the spin contribution is evident in the temperature region in which heat capacity measurements have been made (cf. p. 876). When the nuclear spin entropies are omitted it is, however, necessary to retain a "symmetry number," viz., 2 for a symmetrical and 1 for an unsymmetrical diatomic molecule. The entropy equations are then, respectively,

$$(i) \quad S_{\text{rot.}}^0 = R \ln IT + 177.7 - R \ln 2, \quad (174a)$$

$$(ii) \quad S_{\text{rot.}}^0 = R \ln IT + 177.7. \quad (174b)$$

The moment of inertia of hydrogen chloride, from spectroscopic measurements (p. 569), is 2.65×10^{-40} c.g.s. unit, so that $S_{\text{rot.}}^0$ at 298°K . is 8.1 E.U. per mole. The value of $S_{\text{tr.}}^0$ from (167) is 36.7, making a total of 44.8 E.U. per mole, which may be compared with 44.65 obtained by accurate summation [equation (168)] with due allowance for the two isotopic forms; in neither case is the nuclear spin contribution included. The entropy of hydrogen chloride at 298°K . has also been evaluated from thermal measurements based on the third law, and the result is 44.5 E.U. per mole; this is so close to those obtained by statistical methods as to provide strong evidence for the accuracy of the third law of thermodynamics.

With hydrogen chloride, as for most other stable diatomic molecules, it is evident that the vibrational energy contribution to the partition function is negligible, but for some substances, e.g., chlorine (cf. p. 337), this contribution must be included. For most purposes it is sufficiently accurate to use the approximate expression for the vibrational partition function, as given by (283), p. 337; insertion in (155) then gives

$$S_{\text{vib.}}^0 = \frac{R}{e^{hc\omega/kT} - 1} \cdot \frac{hc\omega}{kT} - R \ln (1 - e^{-hc\omega/kT}), \quad (175a)$$

where $\omega \text{ cm.}^{-1}$ is the equilibrium vibration frequency of the diatomic molecule. At moderately low temperatures, especially if ω is relatively large, this expression reduces virtually to zero, as tacitly assumed above. On the other hand, at high temperatures it becomes

$$S_{\text{vib.}}^0 = R \left(1 + \ln \frac{kT}{hc\omega} \right). \quad (175b)$$

It should be noted that equations (169) to (175) hold only for diatomic molecules, and analogous, but more complicated, approximate expressions have been developed for molecules containing three or more atoms. The general equations (167) and (168) are, of course, always applicable, but the number of terms involved in the accurate summation in the latter increases greatly with the complexity of the molecule. The labor has been simplified, however, by the use of various mathematical devices. By means of the foregoing methods the entropies of a large number of diatomic and polyatomic molecules have been calculated; the agreement in the great majority of cases with the values based on the third law of thermodynamics is so good as to establish the reliability of the latter.¹⁵

Apparent Exceptions to Third Law.—There are a number of substances which appear to provide exceptions to the third law; the first of these to be considered is hydrogen. The entropy due to translation is 28.09 E.U., by (167), and the internal entropy *including* the nuclear spin contribution, $R \ln (2i + 1)^2$, where $i = \frac{1}{2}$, that is $R \ln 4$, is 5.89 E.U., making a total of 33.98 E.U. per mole at 298.16° K. The entropy of hydrogen in the standard state, obtained from thermal data, involving measurements of the heat capacities of solid, liquid and gaseous forms, is 29.64 E.U., based on the third law assumption. The discrepancy between the two values must be attributed to the fact that the rotational levels which characterize the ortho- and para-molecules are carried over into the solid state, so that there is still some randomness of distribution even at the lowest temperatures at which heat capacity measurements have been made. The entropy of the solid, derived from such data, cannot be zero. Taking into account the fact that the hydrogen employed contained the ortho- and para-forms in the normal ratio of 3 to 1, because of the reluctance of nuclear spins to be reversed (p. 94), it can be calculated that the additional entropy should be 4.39 E.U., which must be added to the third law value, giving a total of 34.04 E.U. per mole, in good agreement with the statistical value. This result shows, incidentally, that in the case of hydrogen the thermal entropy also includes the nuclear spin contribution, and

the "practical" or virtual entropy to be used in conjunction with other third law measurements, should have this quantity, i.e., $R \ln 4 = 2.75$, subtracted; the virtual entropy is, therefore, $33.98 - 2.75 = 31.23$ e.u. per mole at 298.16°K .

The existence of rotation in solid hydrogen makes the behavior of this substance provide a definite exception to the third law of thermodynamics. Exceptions of this type are, however, very rare, although in addition to hydrogen, deuterium and possibly one or two of their compounds behave in an analogous manner. The entropy of water obtained by the statistical method is about 0.95 e.u. higher than the third law value; this may be due to rotation of the molecule in the solid, although the suggestion has been made that the entropy is not zero at 0°K , because of the uncertainty of the position of some hydrogen bonds in the crystal.

Another type of exceptional behavior is provided by carbon monoxide, nitric oxide and nitrous oxide; the statistical entropies are all about 1 e.u. higher than the third law values, so that these substances do not have zero entropy in the solid state at 0°K . It is improbable that in these instances there is rotation in the solid at very low temperatures, and the interpretation generally accepted is that because of the similarity of the two ends of the molecule, the crystal lattice is unable to distinguish between the alternative arrangements

CO and OC, NO and ON, NNO and ONN.

Instead of all the molecules being oriented in one direction in the crystal, as would be necessary for zero entropy, there is to some extent a random distribution leading to a finite entropy at 0°K . The maximum value, assuming a completely random arrangement in the crystal, would be $R \ln 2$, i.e., 1.38 e.u., so that the orientation in the solid is at least partially ordered as regards the direction of the molecules at low temperatures.

Restricted Internal Rotation.—An apparent discrepancy between the entropy of ethane as derived from partition functions and the value based on the third law of thermodynamics has brought to light some interesting facts concerning rotation about single bonds. In evaluating the entropy of ethane it was assumed that the two methyl groups rotated freely with respect to each other about the C—C axis, and an appropriate factor was included in the partition function. The value obtained in this manner was 56.36 e.u. per mole at 25°C ., whereas the third law entropy was known to be 54.85 e.u. per mole. Analogous discrepancies have also been found with other molecules in which free internal rotation of two groups about a single bond had been expected, e.g., propane, tetramethylmethane, *n*- and iso-butane, methyl and ethyl alcohols, methyl and dimethylamine, and acetone. In similar compounds, in which the question of rotation of two groups with respect to one another does not arise, e.g., methyl chloride and bromide, methane, etc., the entropies obtained by the two methods were in excellent agreement. From these and various other considerations it has been concluded that, in general, internal rotation of certain groups about a single bond is not completely free, but is restricted to some extent. This view concerning restricted internal rotation does not contradict the organic chemist's concept of "free" rotation, insofar as isomeric structures with different configurations do not exist. The reason why the two ideas do not conflict is that the energy barriers which prevent completely free rotation of two methyl groups, or of a methyl and a

hydroxyl or amino group, with respect to each other are of the order of 3 kcal. This means that the rate at which one configuration changes into another is very rapid at ordinary temperatures. The energy barriers restricting internal rotation would have to be at least 20 kcal. if the different configurations were to be stable and isomeric forms were to be capable of existing under normal conditions.

By making certain assumptions concerning the form of the potential barrier restricting internal rotation, it has been possible to develop various complicated expressions for the appropriate partition functions. However, the use of these expressions requires a knowledge of the maximum height of the potential barrier, and apart from an indication which may be obtained from spectroscopic data, there is no independent method for evaluating this quantity. The procedure adopted at the present time, therefore, is to calculate the entropy, or other thermodynamic property, on the assumption that there is completely free internal rotation. From the difference between this result and that obtained by direct measurement, it is possible to calculate the height of the restricting potential barrier. In this manner the following values were obtained: 2.75 kcal. for ethane, 3.40 kcal. for propane, 3.00 kcal. for methylamine, 3.40 kcal. for methanol, and 1.00 kcal. for acetone. The restricting potential in dimethylacetylene is quite small, less than 0.5 kcal., and this suggests that it is repulsion between hydrogen atoms on the rotating groups which is the main cause of the restriction to free rotation. The height of the barrier having thus been obtained from one thermodynamic property, it can be utilized for the calculation of others.¹⁶

Entropy Data.—In Table 148 are given some of the most recent values for the entropies of gases, excluding nuclear spin effects, but including any contribution resulting from partial random distribution retained in

TABLE 148. MOLAR ENTROPIES OF GASES IN THE STANDARD STATE AT 25° C.

Hydrogen	31.23 E.U.	Nitric oxide	50.34 E.U.	Carbon dioxide	51.07 E.U.
Nitrogen	45.79	Water	45.17	Ammonia	45.9
Oxygen	49.03	Nitrous oxide	52.58	Methane	44.4
Hydrogen chloride	44.64	Deuterium	34.36	Ethane	54.9
Hydrogen bromide	47.48	Chlorine	53.31	Ethylene	52.5
Hydrogen iodide	49.4	Bromine	58.67	Propane	64.7
Carbon monoxide	47.32	Iodine	62.29	Sulfur dioxide	59.2

the crystal at low temperatures; these are the data to be used for calculations of free energy. For simple molecules the entropies recorded are those obtained from the partition functions, but where there is a possibility of restricted internal rotation the values based on the third law are given.

It is of interest to calculate with these data the entropy change accompanying the familiar "water gas" reaction at 25° c.

$$\begin{aligned}\text{CO}_2(g) + \text{H}_2(g) &= \text{CO}(g) + \text{H}_2\text{O}(g), \\ \Delta S^0 &= (47.32 + 45.17) - (51.07) + 31.23 = 10.19 \text{ E.U.}; \\ \therefore \Delta F^0 - \Delta H^0 &= -T\Delta S^0 = -3,037 \text{ cal. at } 298.16^\circ \text{ K.}\end{aligned}$$

Since ΔH° at 298.16°K . is 9,840 cal., it follows that ΔF° is 6,800 cal., which may be compared with the figure 7,200 cal. obtained from direct measurement of equilibria (p. 836). If the value of ΔF° is required at any other temperature, then use may be made of (65), provided the heat capacity constants α , β , γ , etc. are known over the appropriate temperature range. Inserting these values and also that for ΔF° at 298.16°K . it is possible to calculate the constant I ; once this is known ΔF° , or K_p , may be evaluated at any desired temperature within the range for which the heat capacity data hold.

By combining the standard entropies for gases with those for solids and liquids, based on the third law assumption, as given in Table 147, the free energy changes accompanying heterogeneous processes may be estimated in a manner exactly similar to that described for a homogeneous gas reaction.

Free Energy and Equilibrium Constants from Partition Functions.—It is possible by a combination of the statistical methods of deriving free energies and heat capacities, to calculate the former directly at any desired temperature without evaluating the entropy as an intermediate step. The equation (113), p. 230, may be written

$$F^\circ = E^\circ + PV - TS^\circ \quad (176)$$

for a substance in its standard state; in this state a gas is assumed to be ideal, and so it follows that, since $PV = RT$ for 1 mole of gas,

$$F^\circ = E^\circ + RT - TS^\circ. \quad (177)$$

The energy E° is made up of (a) the energy of the lowest (zero-point) level, which may be represented by E_0° , referred to an arbitrary energy zero that is the same for all substances; (b) the translational energy, $\frac{3}{2}RT$; and (c) the energy due to internal degrees of freedom, $E_{\text{int.}}^\circ$, obtained from the appropriate form of (274), page 333; thus,

$$E^\circ = E_0^\circ + \frac{3}{2}RT + RT^2(d \ln Q_{\text{int.}}/dT), \quad (178)$$

where $Q_{\text{int.}}$ refers to internal energy only. Since $S^\circ = S_{\text{tr.}}^\circ + S_{\text{int.}}^\circ$, the use of (155) and (167) gives

$$S^\circ = \frac{3}{2}R \ln M + \frac{5}{2}R \ln T - 2.311 + R[T(d \ln Q_{\text{int.}}/dT) + \ln Q_{\text{int.}}]. \quad (179)$$

Substituting the values of E° and S° from (178) and (179) in (177), it can be readily shown that

$$(F^\circ - E_0^\circ)/T = -\frac{3}{2}R \ln M - \frac{5}{2}R \ln T - R \ln Q_{\text{int.}} + 7.277. \quad (180)$$

It is therefore possible to determine $(F^\circ - E_0^\circ)/T$, called the **free energy function**,* for gases at various temperatures provided the necessary data are available for the evaluation of the partition functions. The values for a number of substances have been tabulated in the literature for temperatures up to 3000°K . or more. From these tables the quantity $\Delta(F^\circ - E_0^\circ)/T$ for a

* The use of this term is justified by the fact that the internal energy E_0° is identical with the free energy at the absolute zero.

chemical reaction can be estimated, provided the reactants and products have been studied in the manner described. By definition (p. 827) $-\Delta F^0 = RT \ln K_p$, and since

$$\Delta F^0 = \Delta(F^0 - E_0^0) + \Delta E_0^0, \quad (181)$$

it follows that

$$-\frac{\Delta F^0}{T} = R \ln K_p = -\Delta \left(\frac{F^0 - E_0^0}{T} \right) - \frac{\Delta E_0^0}{T}. \quad (182)$$

In order to determine K_p or ΔF^0 it is still necessary to know ΔE_0^0 , which is the energy change accompanying the reaction at the absolute zero. This may be obtained from the spectroscopic heats of dissociation (pp. 575, 1165), which refer to the lowest energy levels, by means of (178), utilizing heats of reaction at constant volume, which give ΔE^0 , after correction to the standard states, or in other ways. With ΔE_0^0 available and the tables of the free energy functions, the equilibrium constant or free energy change for a reaction may be ascertained. The data in Table 149, for example, permit the values of K_p

TABLE 149. FREE ENERGY FUNCTIONS AND THE WATER GAS EQUILIBRIUM

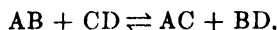
Temp.	CO ₂	$-(F^0 - E_0^0)/T$			$-\Delta(F^0 - E_0^0)/T$	K_p	
		H ₂	CO	H ₂ O		calc.	obs.
900° K.	53.074	32.020	48.114	46.106	9.126	0.451	0.46
1000	54.137	32.752	48.876	46.999	8.986	0.719	0.71
1100	55.125	33.417	49.570	47.820	8.848	1.04	1.01
1200	56.049	34.027	50.210	48.579	8.712	1.41	1.37
1300	56.917	34.589	50.804	49.289	8.587	1.80	1.78

for the water gas reaction at various temperatures to be evaluated; ΔE_0^0 is estimated from thermal measurements to be 9,640 cal. The agreement between observed and calculated values of the dissociation constant in this and other cases is very striking. The significance of the achievement is only fully realized when it is stated that in some cases *all* the data required for the evaluation of free energies, at any desired temperature, can be obtained from spectroscopic measurements.

By combine (182) with (180) it is possible to derive

$$-\ln K_p = \frac{\Delta E_0^0}{RT} - \frac{3}{2} \sum \nu \ln M - \frac{5}{2} \nu \ln T - \sum \nu \ln Q_{\text{int.}} + \sum \nu \frac{7.277}{R} \quad (183)$$

for any reaction. This may be regarded as an alternative method for calculating K_p when tabulated free energy functions are not available; its chief use is, however, in an approximate form for reactions involving no change in the number of molecules, viz.,



for then (183) becomes

$$-\ln K = \frac{\Delta E_0^0}{RT} - \frac{3}{2} \ln \frac{M_{AC} \times M_{BD}}{M_{AB} \times M_{CD}} - \ln \frac{Q_{AC} \times Q_{BD}}{Q_{AB} \times Q_{CD}}, \quad (184)$$

where K may be K_p , K_c or K_x since they are identical for this type of reaction. Provided the temperature is not too high, so that rotational energy levels only

need be considered, $Q_{\text{int.}}$ may be put equal to $Q_{\text{rot.}}$, and this, excluding the nuclear spin factor, may be written

$$Q_{\text{rot.}} = 8\pi^2 I k T / h^2 s, \quad (185)$$

where s is the symmetry number (p. 875); (184) then becomes

$$-\ln K = \frac{\Delta E_0^0}{RT} - \frac{3}{2} \ln \frac{M_{AC} \times M_{BD}}{M_{AB} \times M_{CD}} - \ln \frac{I_{AC} \times I_{BD}}{I_{AB} \times I_{CD}} + \ln \frac{s_{AC} \times s_{BD}}{s_{AB} \times s_{CD}}. \quad (186)$$

This equation is reasonably accurate, since the vibrational contributions to the partition function, which have been ignored, would probably cancel each other almost exactly in any case. In recent years (186) has found particular application in the study of isotopic exchange reactions. Consider, for example, the reaction



then

$$-\ln K = \frac{\Delta E_0^0}{RT} - \frac{3}{2} \ln \frac{M_{\text{D}_2} \times M_{\text{HI}}^2}{M_{\text{H}_2} \times M_{\text{DI}}^2} - \ln \frac{I_{\text{D}_2} \times I_{\text{HI}}^2}{I_{\text{H}_2} \times I_{\text{DI}}^2}, \quad (187)$$

the symmetry numbers being 2 for hydrogen and deuterium, and 1 for the two iodides. Since ΔE_0^0 is known to be 119 cal., and the moments of inertia of the four molecular species can be obtained from spectroscopic measurements, $\ln K$ can be evaluated at any temperature. If the spectroscopic data for the two isotopic forms of each molecule are not available, the moment of inertia of the compound containing the less common isotope may be calculated from that of the more abundant form; the assumption is made that the interatomic distances are the same in each case.

The equations given above apply to reactions involving diatomic molecules, but analogous relationships may be derived for molecules containing a larger number of atoms. The equations are more complicated, however, as three different moments of inertia may be involved.¹⁷

An alternative method of deriving the equilibrium constant from partition functions, which is useful in some connections (see p. 1100), is as follows. The total energy E , with reference to an arbitrary energy zero, is given by (274), p. 333, as

$$E = E_0 + RT^2(\partial \ln Q / \partial T)_V,$$

where Q is the complete partition function, including translational and internal contributions. Further, utilizing the Stirling approximation, $\ln N! \approx N \ln N - N$, the equation (153) for the total entropy becomes

$$S = RT(\partial \ln Q / \partial T)_V + R \ln (Q/N) + R,$$

and if these expressions are introduced into the thermodynamic equation, $F = E + RT - TS$, it is seen that the molar free energy is

$$F = E_0 - RT \ln (Q/N). \quad (188)$$

The free energy change for a gas reaction may then be represented by

$$\Delta F = \Delta E_0 - RT \sum \nu \ln (Q/N), \quad (189)$$

where $\Sigma \nu \ln (Q/N)$ is defined by (cf. p. 824)

$$\Sigma \nu \ln (Q/N) = \ln \frac{(Q/N)_L^L (Q/N)_M^M \cdots}{(Q/N)_A^A (Q/N)_B^B \cdots} = \ln \frac{(Q/N)_{\text{products}}^{\nu}}{(Q/N)_{\text{reactants}}^{\nu}}, \quad (190)$$

for the general reaction considered earlier in the chapter (p. 818). If the reactants and products are all in their standard states, (189) becomes

$$\Delta F^0 = \Delta E_0^0 - RT \Sigma \nu \ln (Q^0/N),$$

and since ΔF^0 is equivalent to $-RT \ln K$, it follows that

$$K = \frac{(Q^0/N)_{\text{products}}^{\nu}}{(Q^0/N)_{\text{reactants}}^{\nu}} e^{-\Delta E_0^0/RT}. \quad (191)$$

It will be noted that the nature of the standard state has not been defined. If it is chosen as the ideal gas at 1 atm. pressure, K is then K_p , and the translational portion of Q^0 , which is the only contribution that is affected, refers to 1 atm. pressure. Another standard state, which will be used in Chapter XIII, is represented by a concentration of 1 molecule per cc.; K is then K_c , and Q_{tr}^0 is now given by (164) with V equal to unity, i.e., $(2\pi mkT)^{1/2}/h^3$, while at the same time the N factors are omitted from (191). The contributions of the internal, i.e., rotational and vibrational, degrees of freedom are unaffected by the choice of standard state.

REFERENCE

- Butler, J. A. V. "Fundamentals of Chemical Thermodynamics," Vols. I and II.
 Guggenheim, E. A. "Modern Thermodynamics."
 Lewis, G. N. and Randall, M. "Thermodynamics, etc."
 MacDougall, F. H. "Thermodynamics and Chemistry"; "Physical Chemistry."
 Nernst, W. "The New Heat Theorem."
 Parks, G. S. and Huffman, H. M. "The Free Energies of Some Organic Compounds."
 Pease, R. N. "Equilibrium and Kinetics of Gas Reactions."
 Taylor, H. S. and Glasstone, S. "Treatise on Physical Chemistry," 3rd. ed., Vol. I, Chap. IV (J. G. Aston).
 van't Hoff, J. H. "Studies in Chemical Dynamics."
1. Le Chatelier, *Compt. rend.*, **100**, 441 (1885).
 2. Bodenstein, *Z. phys. Chem.*, **13**, 56 (1894); **22**, 1 (1897); Bodenstein and Pohl, *Z. Elek.*, **11**, 373 (1905); Langmuir, *J. Am. Chem. Soc.*, **28**, 1357 (1906); Nernst and von Wartenberg, *Z. anorg. Chem.*, **49**, 213 (1906); *Z. phys. Chem.*, **56**, 526; Löwenstein, *ibid.*, **54**, 707 (1906); Haber and Le Rossignol, *Z. Elek.*, **14**, 181 (1908); Trautz, *ibid.*, p. 534; Simons and Hildebrand, *J. Am. Chem. Soc.*, **46**, 2183 (1924); Taylor and Crist, *ibid.*, **63**, 1377 (1941).
 3. Verhoeck and Daniels, *ibid.*, **53**, 1250 (1931); Beeson and Yost, *ibid.*, **61**, 1432 (1939); *J. Chem. Phys.*, **7**, 44 (1939); Spencer, *ibid.*, **8**, 503 (1940); Perlman and Rollefson, *ibid.*, **9**, 362 (1941).
 4. Larson and Dodge, *J. Am. Chem. Soc.*, **45**, 2198 (1923); **46**, 367 (1924).
 5. Edgar and Schuyler, *ibid.*, p. 64; Trimble and Richardson, *ibid.*, **62**, 1018 (1940); Halford and Brundage, *ibid.*, **64**, 36 (1942).
 6. DuBois, *J. Chem. Ed.*, **14**, 324 (1937); Bent and French, *J. Am. Chem. Soc.*, **63**, 568 (1941).
 7. Dimroth, *Annalen*, **377**, 127 (1910).
 8. Smyth and Adams, *J. Am. Chem. Soc.*, **45**, 1167 (1923); Bäckström, *ibid.*, **47**, 2446 (1925); Southard and Royster, *J. Phys. Chem.*, **40**, 435 (1936).

9. Isambert, *Compt. rend.* **92**, 919 (1881); **93**, 731; **94**, 958 (1882); **97**, 1212 (1883); Naumann, *Ber.*, **18**, 1157 (1885); Walker and Lumsden, *J. Chem. Soc.*, **71**, 428 (1897); Briggs, *J. Phys. Chem.*, **28**, 1121 (1924).
10. Preuner, *Z. phys. Chem.*, **47**, 385 (1904); Eastman and Evans, *J. Am. Chem. Soc.*, **46**, 888 (1924); Emmett and Schultz, *ibid.*, **52**, 1782 (1930); **55**, 1376, 1390 (1933).
11. Richards, *Z. phys. Chem.*, **42**, 129 (1902); Nernst, *op. cit.*
12. Einstein, *Ann. Phys.* **22**, 180 (1907); Planck, *Ber.*, **45**, 5 (1912); Lewis and Gibson, *J. Am. Chem. Soc.*, **39**, 2554 (1917); **44**, 1008 (1922).
13. Gerke, *ibid.*, p. 1684.
13. Tetrode, *Ann. Phys.*, **38**, 434 (1912); **39**, 255 (1913), Sackur, *ibid.*, **40**, 67 (1913); Stern, *Phys. Z.*, **14**, 629 (1913).
15. For reviews, see Rodebush, *Chem. Rev.*, **9**, 319 (1931); Kassel, *ibid.*, **18**, 277 (1936); Wilson, *ibid.*, **27**, 17 (1940); Glasstone, *Ann. Rep. Chem. Soc.*, **32**, 66 (1935); "Theoretical Chemistry," Chap. VIII; Thompson, *Ann. Rep. Chem. Soc.*, **38**, 54 (1941).
16. Kemp and Pitzer, *J. Chem. Phys.*, **4**, 749 (1936); *J. Am. Chem. Soc.*, **59**, 276 (1937); Pitzer and Gwinn, *J. Chem. Phys.*, **10**, 428 (1942); for reviews, see Thompson, ref. (15); Glasstone, *op. cit.*, ref. (15).
17. Urey and Greiff, *J. Am. Chem. Soc.*, **57**, 321 (1935); Farkas, "Orthohydrogen, Parahydrogen and Heavy Hydrogen."

For numerous references to work on entropy and allied subjects the reader should consult the issues of the Journal of the American Chemical Society and the Journal of Chemical Physics for recent years.

CHAPTER XII

ELECTROCHEMISTRY

ELECTROLYTIC CONDUCTION

Electrolytic Conductors.—For the subject of physical chemistry the most important conductors are those of the electrolytic type, i.e., **electrolytes**; they are distinguished from electronic conductors, such as metals, by the fact that the passage of an electric current is accompanied by the transfer of matter. There are two main groups of electrolytic conductors; the first consists of pure substances, e.g., fused salts, and the second of solutions. The most thoroughly studied examples of the latter are solutions of acids, bases and salts in water. A few solids act as mixed conductors, behaving in a manner intermediate between those of the electronic and electrolytic types; for instance, β -silver sulfide conducts to the extent of 80 per cent electrolytically, the remainder being electronic. Solutions of alkali metals in liquid ammonia and certain molten alloys also conduct by both mechanisms simultaneously.

The Phenomena of Electrolysis.—When an electric current is passed through an electrolytic conductor, the transport of matter becomes manifest at discontinuities in the conducting system. For example, if two wires, preferably of platinum, attached to the two poles of a voltaic battery acting as a source of current, are immersed in a dilute aqueous solution of an acid, bubbles of hydrogen and oxygen, respectively, are evolved at the wires, as was first observed by W. Nicholson and A. Carlisle (1800). If the electrolytic solution had contained a salt of copper or silver, then the corresponding metal would have been liberated instead of hydrogen. The phenomena associated with electrolysis were studied by M. Faraday (1832–33), and the nomenclature which he used, and which is still employed, was devised for him by W. Whewell. The electrodes, that is, the wires or plates leading the current into and out of the electrolytic solution, are called the **anode** and **cathode**, respectively; the former is the one attached to what is referred to by convention as the “positive” pole of the battery while the latter is connected to the “negative” pole. In the course of electrolysis hydrogen gas or a metal is set free at the cathode, and oxygen or a halogen appears at the anode, provided the electrode material is not attacked. If the anode consists of zinc, copper, iron or other base metal it will tend to dissolve under the influence of the current. To account for the observations made during the passage of current through an electrolyte, Faraday assumed the flow of electricity to be associated with the movement of charged particles, although he did not speculate as to their origin; these were called **ions** (Greek: *wanderer*),

the particles carrying a positive charge and moving in the direction of the electric current, i.e., towards the cathode, were referred to as **cations**, and those having a negative charge, and moving towards the anode, were called **anions** (Fig. 228). On reaching the electrodes these ions were assumed to have their charges neutralized, and to be deposited in the form of normal atoms or molecules.

Faraday's Laws of Electrolysis.—The relationship between the quantity of electricity passing through an electrolyte and the amount of material liberated at the electrodes was discovered by Faraday; his results may be expressed in the form of the two laws of electrolysis.

(i) *The amount of chemical decomposition produced by a current, that is the amount of any substance deposited or dissolved, is proportional to the quantity of electricity passed.*

(ii) *The amounts of different substances deposited or dissolved by the same quantity of electricity are proportional to their chemical equivalent weights.*

The laws can be tested by passing the same current simultaneously through a series of solutions, e.g., dilute sulfuric acid, silver nitrate and copper sulfate, when the amounts of material deposited on the cathodes, i.e., hydrogen, silver and copper, respectively, will be proportional to the respective equivalent weights, viz., 1 to 107.88 to 31.78; this verifies the second law. The amounts of silver or copper dissolved from anodes of these metals are found to be the same as those deposited on the cathode. The first law may be confirmed by passing different currents for various periods of time through a given solution and determining the amount of material deposited on the cathode or dissolved from the anode. The quantity of electricity is equal to the product of the current strength and the time during which it flows. In this manner the reliability of Faraday's laws has been established for a wide variety of conditions, provided the conduction is entirely electrolytic in nature.

The Ampere and the Faraday.—The practical unit of current is the **ampere**; the absolute ampere is one-tenth of the e.m. (c.g.s.) unit, described on p. 5, but this differs slightly from the international ampere,* defined as the current which flowing for 1 sec. will cause the deposition of 1.11800 mg. of silver from the solution of a silver salt. The unit quantity of electricity is the **coulomb**, i.e., 1 amp. sec., and hence 1 coulomb deposits 1.11800 mg. of silver, i.e., 0.001118/107.88 g. equiv. It follows, therefore, that 96,494 coulombs would be required to deposit

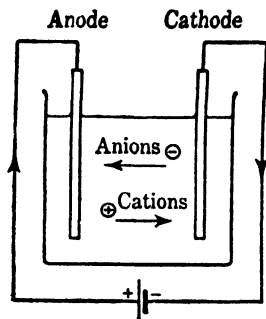


FIG. 228. Passage of current through an electrolyte

* This is actually 0.9999 absolute amp.

or dissolve 1 g. equiv. of silver, and hence, by Faraday's second law, 1 g. equiv. of any other element. In view of the uncertainties, connected with the matter of units, the presence of impurities in the silver, and the question of chemical and physical atomic weights, it is sufficiently accurate to state that the passage of 96,500 international coulombs* of electricity results in the deposition of 1 g. equiv. of any element. This quantity is called the **faraday**, and is given the symbol F ; as seen in Chapter I it is believed to be the amount of electricity carried by 1 g. equiv. of any ion, gaseous or in solution, having a unit charge. If z is the valence of an atom or radical, then the corresponding g. ion contains z g. equiv. and hence must carry a charge of zF coulombs. In general, the number of unit charges carried by any ion is equal to its valence.

Measurement of Quantity of Electricity.—If a current of I amp. flows for t sec., the number of coulombs passing is It ; hence by Faraday's first law the amount of any material deposited on an electrode will be $It/96,500$ g. equiv. If e is the equivalent weight of the substance, then the total weight deposited will be $It e/96,500$ g.; this formula may be regarded as the mathematical expression of the laws of electrolysis, and it may be utilized for the exact measurement of quantity of electricity, or of current strength. The apparatus used for the purpose is known as a **voltameter**, or better as a **coulometer** (T. W. Richards, 1902). The most accurate is probably the silver coulometer, in which a solution of silver nitrate is electrolyzed under carefully controlled conditions. The silver deposited on the cathode, generally consisting of a platinum basin containing the electrolyte, is washed, dried and weighed; from the weight, the quantity of electricity passing, i.e., It , may be estimated by the formula given above. If the time t is known then the current strength I , assuming it to have remained constant, can be evaluated. For many laboratory purposes the copper coulometer is employed; it contains an acidified solution of copper sulfate, together with a small amount of alcohol or tartaric acid to prevent oxidation of the deposited copper. The quantity of electricity is determined from the fact that 1 coulomb deposits 0.3294 mg. of copper. For rough measurements the electrolytic gas coulometer may be used; it is conveniently set up with nickel electrodes in an alkaline solution, and the total volume of hydrogen and oxygen is measured. One faraday of electricity liberates 16.8 liters of gas at s.t.p.¹

Significance of Faraday's Laws.—It was seen on p. 1 that the fact that one faraday of electricity is required to liberate a g. equiv. of any element leads to the concept of the atomic structure of electricity, with the electron as the unit. A further consequence of these arguments is that a univalent cation is formed by the loss of a single electron from an atom, e.g., $\text{Na} \rightarrow \text{Na}^+ + e$; similarly, a bivalent cation results from the loss of two electrons, e.g., $\text{Zn} \rightarrow \text{Zn}^{++} + 2e$, and so on.† In general, an ion

* This is equivalent to 96,490 absolute coulombs.

† In order to avoid confusion with the base of natural logarithms, the symbol e is used for the electronic charge in this chapter.

carries the same number of charges as its valence, and it differs from the corresponding uncharged particle by this number of electrons. These ideas are, of course, tacitly involved in the theory of electrovalence developed in Chapter I.

As a consequence of Faraday's laws it is required that the passage of a given quantity of electricity should result in the discharge (or formation) of equivalent amounts of different ions at anode and cathode. Hence, when a solution is electrolyzed the number of electrons taken up by the cations at the cathode must be equal to the number released simultaneously at the anode. This will presumably be also equal to the electrons which pass through the *external* connection from the anode to the cathode, i.e., in the direction opposite to the conventional flow of positive current. It is thus possible to obtain a simple picture of electrolysis in terms of electron transfer.

The Mechanism of Electrolytic Conduction.—The first attempt to explain electrolysis was made by T. von Grotthuss (1805); he maintained that the electric field orients the molecules of electrolyte in chains, so that the negative parts point to the positive electrode and *vice versa*. The electrodes then attract the ends of the chains, thus liberating the positive part of one end molecule at the negative electrode, and the negative part of the other end molecule at the positive electrode. The residual parts of the end molecules are then supposed to exchange partners with adjacent molecules, and this interchange proceeds through the whole length of the chain until a complete set of new molecules is formed. These are again oriented by the current, split up, and so on. Objection was taken to the theory by R. Clausius (1857), chiefly on the ground that only very small expenditure of energy is required for electrolysis in many cases, whereas the views of Grotthuss implied the necessity of a large force to break a molecule into its component parts. To overcome the difficulty Clausius suggested that the positive and negative parts of a molecule were not firmly connected, but were capable of separating from each other for short intervals; these charged particles, or ions, were supposed to carry the current during the short periods of their free existence. According to Clausius, therefore, only a very small number of molecules would be split up into ions at any given instant, but as these ions were removed by discharge at the electrodes, or by recombination, they would be replaced by fresh ions resulting from vigorous vibration or from collisions between molecules. Following on van't Hoff's discovery that electrolytic solutions showed very large deviations from his osmotic pressure equation (p. 663), even for dilute solutions, S. Arrhenius (1887) put forward the **theory of electrolytic dissociation**, as a more explicit form of one he had proposed in 1883, which forms the basis of the modern treatment of electrolytes. The assumption made was that when an acid, base or salt is dissolved in water, a *considerable portion* becomes *spontaneously* dissociated into positive and negative ions; that is,



These ions are free to move independently, and are simply guided to the electrodes of opposite sign by the applied electric field. The proportion of molecules which dissociate into ions was believed to vary with the concentration, and the "degree of dissociation," i.e., the fraction of the total electrolyte split up into ions, was considered to approach unity at infinite dilution; that is to say, in extremely dilute solutions almost the whole of the acid, base or salt is dissociated into ions.

Suppose each molecule of electrolyte splits up into ν ions; then if a solution contains n molecules in a given volume, and α is the degree of dissociation, there will be present in solution $n(1 - \alpha)$ undissociated molecules and $\nu\alpha n$ ions, i.e., a total of $n(1 - \alpha + \nu\alpha)$ particles. If van't Hoff's views concerning osmotic pressure are assumed to hold, then this pressure, like gas pressure, will be proportional to the number of individual particles in a given volume, irrespective of their nature. Had there been no dissociation, the osmotic pressure of the solution would have been proportional to n , but since there are actually $n(1 - \alpha + \nu\alpha)$ particles in the given volume, it follows that the van't Hoff factor i (see p. 674), which gives the ratio of the observed to the calculated osmotic pressure, will be given by

$$i = \frac{n(1 - \alpha + \nu\alpha)}{n} = 1 - \alpha + \nu\alpha; \quad (1)$$

$$\therefore \alpha = (i - 1)/(\nu - 1). \quad (2)$$

The value of i for any solution can be calculated, approximately, from any of the colligative properties described in Chapter IX; Arrhenius used the ratio of observed to calculated freezing point depressions, and by means of (2) he calculated what he claimed to be the degrees of dissociation of various electrolytes at different concentrations. As will be explained shortly, he also derived α from the electrical conductance of the same solutions, and the agreement between the two sets of values was regarded as very strong evidence for the theory of electrolytic dissociation. It is now believed that both methods of calculating α were based on unjustifiable assumptions, and so the agreement is mainly fortuitous; nevertheless, the view that acids, bases and salts split up spontaneously into ions to a very large extent when dissolved in water, or certain other solvents, is almost universally accepted. The chief modification made in recent years (p. 956) is that most salts and strong acids and bases are believed to be almost completely split up into ions, so that the degrees of dissociation are larger than those proposed by Arrhenius for solutions of appreciable concentration.²

The Conductance of Electrolytes.—Solutions, like metallic conductors, obey Ohm's law, except under such abnormal conditions as very high voltages or with very high frequency currents (p. 908). If an electrical pressure, or electromotive force (E.M.F.), of value E is applied to a conductor of resistance R , then by Ohm's law the current I which flows is

$$I = E/R. \quad (3)$$

The practical unit of resistance is the **ohm**, the international ohm being the resistance at 0° C. of a column of mercury, of uniform cross section, 106.3 cm. long and weighing 14.4521 g. The practical unit of E.M.F., or potential, is the **volt**; the absolute value has been defined on p. 6, but the international volt, which careful comparison has shown to be 1.00034 absolute units, is the potential required to force a current of 1 int. amp. through a resistance of 1 int. ohm. It is thus possible to write Ohm's law in the form: amperes = volts/ohms, the prefix "international" being dropped.

Since by (3) the current flowing through a given conductor, under the influence of a constant E.M.F., is inversely proportional to the resistance, the quantity $1/R$ is a measure of the conducting power, and is called the **conductance**; it is expressed in "reciprocal ohms," i.e., ohms^{-1} , sometimes called "mhos." The resistance of any conductor varies directly as its length (l cm.) and inversely as its area (a sq. cm.); that is,

$$R = \rho l/a \text{ ohms}, \quad (4)$$

where ρ is a constant, the **specific resistance**, or **resistivity**, of the conducting material. It is the resistance in ohms of a specimen 1 cm. in length and 1 sq. cm. in cross section. The **specific conductance**, designated by κ , of a given material is defined as $1/\rho \text{ ohms}^{-1} \text{ cm.}^{-1}$; hence (4) may be written

$$R = l/\kappa a. \quad (5)$$

If the conductance of the solution is represented by C , then

$$C = 1/R = \kappa a/l \text{ ohms}^{-1}. \quad (6)$$

The conducting power of all the ions produced by 1 g. equiv. of an electrolyte at any particular concentration may be evaluated by imagining two large parallel electrodes set 1 cm. apart and supposing the *whole* of the solution containing the 1 g. equiv. to be placed between them. The conductance of the resulting system is called the **equivalent conductance** (Λ) at the given concentration. Suppose the 1 g. equiv. of electrolyte is dissolved in v cc. of solution; then this will cover an area of v sq. cm. of the two electrodes, which are 1 cm. apart. It follows, therefore, from (6), since the conductance is Λ when a is v sq. cm. and l is 1 sq. cm., that

$$\Lambda = \kappa v = 1000\kappa/c \text{ ohms}^{-1} \text{ cm.}^2, \quad (7)$$

where c is the concentration of the solution in *g. equiv. per liter*. The equivalent conductance is seen to be numerically equal to the product of the specific conductance and the volume in cc. containing 1 g. equiv., and it is sometimes defined in this way. Some workers have employed the **molar conductance** (μ); it is obtained when κ is multiplied by the volume in cc. containing 1 mole of electrolyte. For uni-univalent electrolytes Λ and μ are identical, since the mole and the g. equiv. are the same.

Measurement of Electrolytic Conductance.—The erratic results obtained in the earliest measurements of the conductance of solutions were due to the

employment of direct current. The gases liberated at the electrodes not only increase the resistance, but they set up a counter E.M.F. of "polarization" (see p. 1013) which opposes the passage of current. The most satisfactory method of overcoming polarization is to employ a rapidly alternating current of low intensity (F. W. Kohlrausch, 1868); the direction of the current is reversed about a thousand times per sec., and the polarization produced by each pulse of the current is neutralized by the next, provided the alternations are symmetrical. The familiar Wheatstone bridge is generally employed to measure the resistance of the solution, a telephone earpiece being used to detect the balance point. The arrangement of the apparatus is shown diagrammatically in Fig. 229; *A* is the cell containing the electrolyte and *B* is a variable standard resistance, e.g., a resistance box, *C* is the source of alternating current, and *D* is the current detector. The resistances of *B*, and of the two arms *lo* and *on*, are adjusted, in the latter case by changing the position of *o*, until no current can be detected; then the resistances in the arms of the bridge are related by

$$A/B = lo/on, \quad (8)$$

provided the resistances of the various leads are negligible. The section *ln* may consist of a uniform graduated wire, or two variable resistance boxes may be used, or two boxes connected by a wire of known resistance for making final adjustments. The resistance of *B* should be of such a magnitude that the bridge is balanced almost at the midpoint of *ln*; any error in setting the position of *o* will then cause least error in the final result. With electrolytes of low conductance it is difficult to obtain a sharp balance point unless the capacity of *A* is compensated by the variable condenser, shown in Fig. 229, connected across

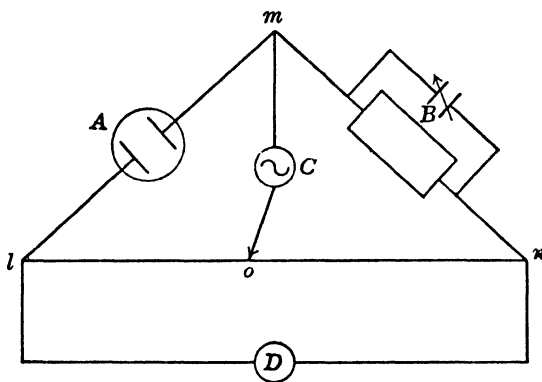


FIG. 229. Measurement of conductance

the box *B*. Since the resistance of *B* is known, and the ratio *lo/on* can be measured, the resistance of the cell *A* is given by (8).

For approximate work the induction coil, introduced by Kohlrausch, may be used as a source of alternating current (A.C.) but, apart from being noisy, it does not give a symmetrical current, and so does not eliminate polarization completely. Various forms of A.C. generators have been recommended, but in modern work electron (vacuum) tube oscillators are almost invariably employed. The telephone receiver, employed by Kohlrausch in 1880, is still

the best form of current detector: it is most sensitive at a frequency of about 1000 cycles, and it should be tuned to the A.C. used. To increase the sensitivity, advantage may be taken of the amplifying properties of the vacuum tube. For measurements in which great accuracy is not required, an A.C. galvanometer or a D.C. instrument, in conjunction with a form of rectifier, may be utilized to detect the balance point of the bridge. Various types of cells have been designed for conductance work; they should be made of insoluble glass or fused silica, and should be carefully washed and steamed before use. In Fig. 230, I, is shown a simple cell often employed in the laboratory,

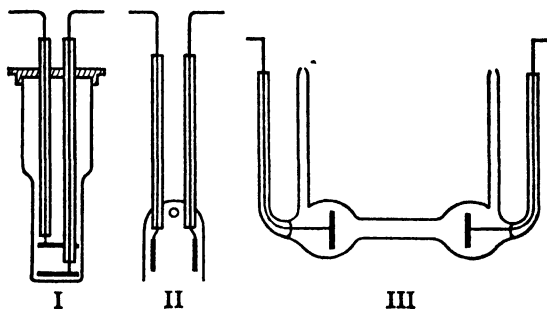


FIG. 230. Cells for measurement of conductance

while a convenient form for industrial purposes is the "dipping cell," II; for accurate work a special type of cell, such as that of G. Jones (1931), Fig. 230, III, must be used. The electrodes are of stout platinum, which do not bend readily, and their positions relative to each other must be fixed. In order to decrease the effect of polarization the electrodes are coated with a layer of finely divided platinum black (F. W. Kohlrausch, 1875); this is best achieved by electrolysis of a 3 per cent solution of chloroplatinic acid containing a little lead acetate. In some cases the platinized electrodes cannot be used, because the finely divided platinum may catalyze the decomposition or oxidation of the solution being studied, or it may adsorb the electrolyte; the use of a gray platinum surface, obtained by heating the platinized electrode to redness, obviates these sources of error. Under certain conditions, especially with high frequency A.C., smooth platinum electrodes will give fairly good results.

Conductance Water.—Specially pure water, known as **conductance water**, should be used for making up the solutions. The purest water, referred to as "ultra-pure," has a specific conductance of about $0.05 \times 10^{-6} \text{ ohm}^{-1} \text{ cm.}^{-1}$ at 18°C. , but it is only required in the most accurate work, and air must be rigidly excluded when it is being employed. For most purposes satisfactory conductance water may be obtained by distilling good distilled water, to which a little permanganate has been added, from a hard glass retort with a block tin or resistance-glass condenser. Water of this type has a specific conductance of $0.8 \times 10^{-6} \text{ ohm}^{-1} \text{ cm.}^{-1}$, mainly due to carbon dioxide dissolved from the air to which it is exposed; it is often called "equilibrium water." Corrections must usually be applied for the conductance of the water; when working with solutions of salts which are not appreciably hydrolyzed (p. 984), the whole of the conductance of the water should be subtracted from that of the measured solu-

tion, but if hydrolysis occurs and the solution is acid this correction is probably *too large*. For an acid solution having a specific conductance greater than about $2 \times 10^{-4} \text{ ohm}^{-1} \text{ cm.}^{-1}$ the water correction may be neglected, since the acid electrolyte represses the ionization of carbonic acid to which the conductance of the equilibrium water is chiefly due. If the conductance of the solution is less than the value given, the correction may be calculated from a knowledge of the dissociation constants (p. 954). The water correction to be applied with alkaline solutions is very uncertain; for most purposes it is better not to subtract the conductance of the water, but for accurate work ultra-pure water should be used.

The Cell Constant.—It is seen from (5) that $\kappa = l/aR$, where l may be taken as the distance between the two electrodes in the conductance cell and a is the area of each. For a given cell l and a are constant, and the quantity l/a is called the **cell constant**. It may be obtained by direct measurement, but this is rarely done, since it is possible to evaluate it by means of a solution of known specific conductance. Potassium chloride solutions are invariably used for this purpose, for they have been measured with great accuracy in cells of known dimensions. A given solution of potassium chloride of specific conductance κ_0 is placed in the experimental cell and its resistance R_0 is measured; then the cell constant K is equal to $\kappa_0 R_0$. If R is the resistance of any other solution in the same cell, then its specific conductance κ is K/R .³

Results.—The results of conductance measurements are usually expressed in the form of the equivalent conductance of the solution, derived from (7), utilizing the known concentration and the value of κ determined experimentally. Some of the data for a few electrolytes of different types are recorded in Table 149, the concentrations being in g. equiv. per liter.

TABLE 149. EQUIVALENT CONDUCTANCES AT 25° IN OHMS⁻¹ CM.²

Concn.	HCl	KCl	NaOH	AgNO ₃	$\frac{1}{2}$ BaCl ₂	$\frac{1}{2}$ NiSO ₄	$\frac{1}{2}$ LaCl ₃	CH ₃ COOH
0.0005	422.74	147.81	246	131.36	135.96	118.7	139.6	—
0.001	421.36	146.95	245	130.51	134.34	113.1	137.0	48.63
0.005	415.80	143.55	240	127.20	128.02	93.2	127.5	22.80
0.01	412.00	141.27	237	124.76	123.94	82.7	121.8	16.20
0.02	407.24	138.34	233	121.41	119.09	72.3	115.3	11.57
0.05	399.09	133.37	227	115.24	111.48	59.2	106.2	7.36
0.10	391.32	128.96	221	109.14	105.19	50.8	99.1	5.20

It is seen that in every case the equivalent conductance, and hence the conducting power of the ions in 1 g. equiv. of electrolyte, increases with decreasing concentration. With hydrochloric and other strong acids, strong bases and most salts, the change in conductance is not very large, but with weak acids and bases, e.g., acetic acid and ammonia, not only does the equivalent conductance fall off rapidly as the concentration is increased, but the values are very low at concentrations of about 0.1 g. equiv. per liter (Fig. 231). Solutions of the former type, i.e., strong acids and bases, and many salts, are called **strong electrolytes** and the latter, mainly organic acids and bases, are known as **weak electrolytes**. It is, of course, impossible to draw a sharp line between the two categories,

ELECTROLYTIC CONDUCTION

for many substances exhibit intermediate behavior, e.g., nickel sulfate; this is particularly evident with nonaqueous solvents, having relatively low dielectric constants.

Equivalent Conductance at Infinite Dilution.—The equivalent conductances of the strong electrolytes in Table 149 are evidently approaching a limiting value as the concentration is diminished; this quantity is called the equivalent conductance at "infinite dilution," and is represented by Λ_0 . It is presumably a measure of the conducting power of 1 g. equiv. of the electrolyte when it is completely split up into ions and

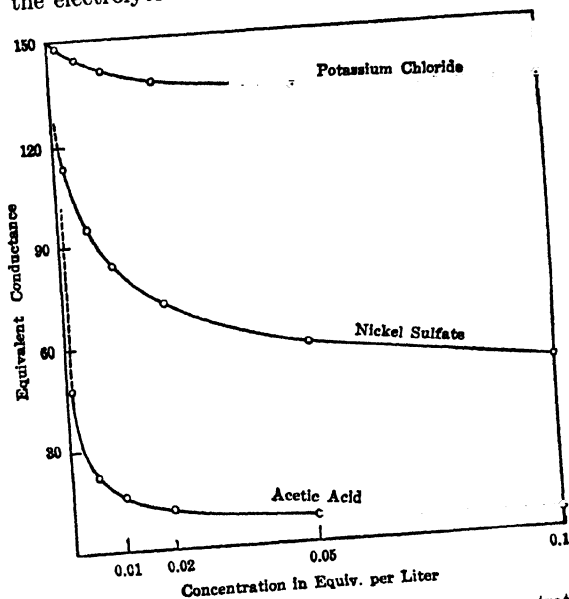


FIG. 231. Variation of equivalent conductance with concentration

the latter are so far apart that they exert no influence on each other. The determination of Λ_0 involves an extrapolation of the measured equivalent conductances in dilute solution to zero concentration; for this purpose Λ should be plotted against a function of the concentration, preferably \sqrt{c} which gives a straight line at high dilutions. Various methods of extrapolation have been employed, but the plot of Λ against \sqrt{c} , proposed by F. W. Kohlrausch (1900), is one of the best. Another procedure for evaluating Λ_0 is to utilize the equation

$$\Lambda_0 = \Lambda + (A + B\Lambda_0) \sqrt{c}, \quad (9)$$

derived from the modern theory of strong electrolytes (p. 905), where A and B , which can be readily calculated, are constants for an electrolyte

of a given valence type in each particular solvent. Since A and B are known, it should be possible to determine Λ_0 from Λ measured at any value of c . Owing to slight departures from theoretical behavior the provisional values of Λ_0 so obtained are not independent of c , but if they are plotted against c and extrapolated as a straight line to zero concentration, the correct value of the equivalent conductance at infinite dilution is obtained (T. Shedlovsky, 1932). The extrapolation formulae cannot be used with weak and intermediate electrolytes, for Λ_0 differs greatly from the measured equivalent conductances in the most dilute solutions, and extrapolation would be very inaccurate. For such electrolytes other procedures, referred to later, are used.

The Independent Migration of Ions.—When the equivalent conductances at infinite dilution of pairs of salts having an ion in common are examined, certain regularities are evident. The data in Table 150, for

TABLE 150. EQUIVALENT CONDUCTANCES AT INFINITE DILUTION AT 18° C.

Electrolyte	Λ_0	Electrolyte	Λ_0	Difference
KCl	130.0	NaCl	108.9	21.1
KNO ₃	126.3	NaNO ₃	105.2	21.1
$\frac{1}{2}$ K ₂ SO ₄	133.0	$\frac{1}{2}$ Na ₂ SO ₄	111.9	21.1

example, are for sodium and potassium salts at 18° c. The difference in conductance of sodium and potassium salts of the same anion is independent of the nature of the latter, and similar results are obtained with other pairs of salts having either anion or cation in common. This phenomenon, discovered by F. W. Kohlrausch (1876), was ascribed by him to the fact that each ion contributes a definite amount to the total conductance of the electrolyte, irrespective of the nature of the other ion; this rule, known as the **law of independent migration of ions**, is strictly true only at infinite dilution when there is no mutual ionic interaction. The law may thus be expressed in the form

$$\Lambda_0 = \lambda_+ + \lambda_-, \quad (10)$$

λ_+ and λ_- being called the **ion conductances**, at infinite dilution, of the cation and anion, respectively; they are the amounts contributed by each ion, independent of the other, to the total equivalent conductance.⁴

From a consideration of the accurate data on the equivalent conductances of strong electrolytes at infinite dilution, in conjunction with transference number measurements (p. 918), it has been possible to allot definite ion conductances to many cations and anions; some of the values are quoted in Table 151, the temperature being 25° c. Although there does not appear to be any connection between the ion conductance and the size of simple ions, for large organic ions and complex inorganic ions the conductances decrease with increasing size towards a limiting value of about 28 ohms⁻¹ cm.² at 25°; this is approximately true for the anions of fatty acids, for alkylammonium ions, and for cobaltammine ions. The

TABLE 151. ION CONDUCTANCES AT INFINITE DILUTION AT 25° IN OHMS⁻¹ CM.²

Cation	λ_+	Anion	λ_-
H ⁺	349.82	OH ⁻	198.5
K ⁺	73.52	Br ⁻	78.4
NH ₄ ⁺	73.4	I ⁻	76.8
Ag ⁺	61.92	Cl ⁻	76.34
Na ⁺	50.11	NO ₃ ⁻	71.44
Li ⁺	38.69	CH ₃ CO ₂ ⁻	40.9
$\frac{1}{2}$ Ba ⁺⁺	63.64	$\frac{1}{2}$ SO ₄ ⁻⁻	79.8
$\frac{1}{2}$ Mg ⁺⁺	53.06	$\frac{1}{2}$ Fe(CN) ₆ ⁻⁻⁻⁻	110.5

equivalent ion conductances of C₆H₁₁CO₂⁻, N(C₂H₇)₄⁺ and Co(NH₃)₆⁺⁺⁺, for example, are about 28, 24 and 26 ohms⁻¹ cm.², respectively.

Ion Conductance and Ionic Speed.—At infinite dilution all the ions that can be derived from an electrolyte are taking part in conducting the current. Hence solutions each containing 1 g equiv. of different electrolytes will possess, under these conditions, *equivalent* numbers of ions; that is, the total charge carried by the ions will be the same for all the solutions. The quantity of electricity that can pass through an electrolyte, and hence its conductance, depends on the product of the number of ions, the charge carried by each ion, and the velocity with which the ions move. Since the total charge is a constant in each case, the equivalent conductance of an electrolyte at infinite dilution can only depend on the speeds of the ions, and it is the difference in the speeds which is responsible for the different values of Λ_0 . If u_+ and u_- are the velocities of cation and anion, respectively, for example under a fall of potential of 1 volt per cm., then the equivalent conductance at infinite dilution must be proportional to the sum of the speeds of the two ions, so that

$$\Lambda_0 = k(u_+ + u_-) = ku_+ + ku_-, \quad (11)$$

where k is the same constant for *all* electrolytes. Since ku_+ represents the contribution of the cation to the total conductance and ku_- that of the anion, it follows on comparison of (10) and (11) that

$$\lambda_+ = ku_+ \quad \text{and} \quad \lambda_- = ku_- \quad (12)$$

The ion conductance is thus proportional to the speed of the ion.

Influence of Temperature and Pressure.—The equivalent conductance at infinite dilution increases with increasing temperature, and a formula of the type

$$\Lambda_{t^\circ} = \Lambda_{25^\circ}[1 + x(t - 25)] \quad (13)$$

is applicable, where Λ_{t° and Λ_{25° are the values at t° and 25° c., respectively, and x is a constant for each electrolyte. For salts x is about 0.019 to 0.021, so that the equivalent conductance at infinite dilution at ordinary temperatures increases by about 1.9 to 2.2 per cent per degree. Acids and bases have slightly smaller temperature coefficients; they are approximately 0.016 and 0.018, respectively. It is evident that an equa-

tion similar to (13) must hold for individual ion conductances, and for nearly all ions, except H^+ and OH^- , the value of α is 0.018 to 0.022; for the hydrogen ion it is 0.0142, and for the hydroxyl ion 0.016. As a rough generalization, it may be said that the higher the ionic conductance the smaller the temperature coefficient, so that the differences between the values for different ions become smaller as the temperature is raised.

The results described apply particularly to infinite dilution, but similar conclusions hold for strong electrolytes at appreciable concentrations. Weak electrolytes, however, show some abnormalities: the conductance of such solutions is dependent mainly on the degree of dissociation, and since this usually decreases with increasing temperature, the conductance may, particularly with concentrated solutions, increase to a maximum and then decrease as the temperature is raised.

The conductance of a solution varies slightly with pressure, the effect being mainly due to the change in the viscosity of the medium and the consequent alteration in the speed, and hence the conductance, of the ions. The viscosity of a dilute solution is generally decreased by an increase of pressure; this is accompanied by an increase of the equivalent conductance, provided the pressure is not too high. The effect is more marked with a weak than with a strong electrolyte, for in the former case increase of pressure tends to favor ionization, but this factor is inappreciable in the latter.

Influence of Solvent.—The results hitherto have referred specifically to solutions in water, and it is necessary to see how far nonaqueous solutions are similar or different. Not all solvents form appreciably conducting systems; benzene and other hydrocarbons give a negligibly small conductance, but nitriles, nitro-compounds, alcohols, ketones, liquid hydrogen cyanide and ammonia yield solutions of salts with a fair conducting power. J. J. Thomson (1893) and W. Nernst (1893) pointed out that in a medium of low dielectric constant the electrostatic attractive force between positive and negative ions will be large, and such solvents would be expected to have a small dissociating influence on an electrolyte; a medium of high dielectric constant, however, should favor dissociation. This connection between dissociation and dielectric constant of the solvent is called the **Nernst-Thomson rule**, which may be regarded as approximately true; a more exact relationship can be derived from the modern theory of electrolytes (p. 907). Solvents of high dielectric constant behave very much like water, the equivalent conductance increasing steadily with decreasing concentration, but in media of low dielectric constant, it is frequently found that there is a minimum in the equivalent conductance at a particular concentration. This concentration $c_{min.}$ is related to the dielectric constant D by the expression $D^3/c_{min.} = \text{constant}$, for a given electrolyte in different solvents (P. Walden, 1922). It is evident that for a solvent of high dielectric constant, e.g., water, the conductance minimum could only be observed at very high concentrations. Under these conditions other factors will be operative which would undoubtedly obscure the minimum. A comprehensive study of the con-

ductance of solutions of tetraisoamylammonium nitrate was made by C. A. Kraus and R. M. Fuoss (1933); the solvents were a series of mixtures of dioxane and water, the dielectric constant varying steadily from 2.2 to 78.6. The conductance minimum was found to move to regions of higher concentration as the dielectric constant of the medium increased; in solvents of very low or very high dielectric constant the minimum occurred at such low or high concentrations, respectively, that it could not be observed. Kraus and Fuoss were able to explain the results and to derive Walden's relationship by assuming that ion triplets, e.g., $+-+$ and $-+-$, were formed, especially in solutions of low dielectric constant (see p. 969).⁵

The equivalent conductance of strong electrolytes at infinite dilution in solvents such as alcohols, nitriles or nitro-derivatives, can be obtained by extrapolation methods similar to those described above. For media of low dielectric constant it is apparent that such procedures are useless. Special methods, based on theoretical equations, can often be employed with success (R. M. Fuoss and C. A. Kraus, 1933 *et seq.*). Once the conductances at infinite dilution of a number of electrolytes have been determined, the ion conductances can be derived in the normal manner; these are additive, like the values in aqueous solution.

Viscosity and Conductance.—If the size of an ion is independent of the nature of the solvent, and Stokes's law (p. 497) may be applied, it follows that the product of the speed of the ion and the viscosity of the medium should be constant. The equivalent conductance of an ion at infinite dilution depends only on its speed (p. 895), and so the product of ion conductance and viscosity should be independent of the nature of the solvent. This relationship, known as **Walden's rule**, was put forward by P. Walden (1906) in the form $\Lambda_0\eta_0 = \text{constant}$ for a given electrolyte, where η_0 is the viscosity of the solvent in which the equivalent conductance at infinite dilution is Λ_0 . Walden claimed that his rule was obeyed by a number of salts, but closer examination has shown that it has certain limitations. Some data for the product of ion conductance and viscosity, i.e., $\lambda\eta_0$, for various ions in different solvents at 25° are given in Table 152. The lack of constancy in the vertical columns shows the

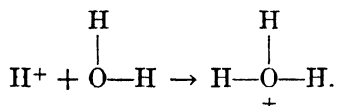
TABLE 152. PRODUCT OF ION CONDUCTANCE AND VISCOSITY AT 25° C.

Solvent	H ⁺	Li ⁺	Na ⁺	NH ₄ ⁺	Cl ⁻	NO ₃ ⁻
Water	3.14	0.36	0.46	0.66	0.69	0.63
Methyl alcohol	0.77	0.22	0.25	0.32	0.29	0.83
Ethyl alcohol	0.64	0.16	0.20	0.20	0.26	0.28
Acetone	0.28	0.22	0.25	0.28	0.33	—
Acetonitrile	—	—	0.24	—	—	—

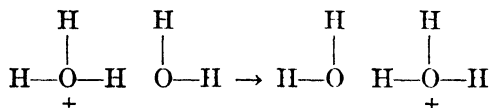
limitations of Walden's rule, and proves that an ion does not have the same effective radius in different solvents. It is probable that in solution ions are **solvated**, that is to say, the ion attaches to itself by physical or

chemical forces, or both, a number of molecules of solvent (cf. p. 920); since the extent of solvation, as well as the size of the solvating molecules, varies with the nature of the solvent, the effective radius will not be constant. The Walden rule appears to be most closely obeyed by ions of the type NR_4^+ , where R is a relatively large radical; a saturated structure of this type will have little tendency to attract solvent molecules, and will, therefore, have the same size in different media.

The values of $\lambda\eta_0$ for ions in water are generally abnormal, but that for the hydrogen ion is particularly so, in agreement with its exceptionally high conductance (Table 151). It was thought at one time that this result could be attributed to the small size of the hydrogen ion or proton, but it now seems certain that the "bare" ion does not exist in water to any measurable extent. The very high affinity of the water molecule for a proton (p. 413) shows, without doubt, that the hydrogen ion in aqueous solution is at least $\text{H}^+(\text{H}_2\text{O})$, or H_3O^+ ; thus,



Analogous ions, e.g., $\text{C}_2\text{H}_5\text{OH}_2^+$, probably exist in alcoholic solutions. The ion H_3O^+ is comparable in size with one of sodium, and so the high conductance cannot be explained by its size. It is now widely accepted that in water, and to some extent also in other hydroxylic media, the conducting power of the hydrogen ion is due to a form of Grotthuss conduction. If a hydrogen ion and a molecule of water approach, e.g.,



it is evident that the transfer of a proton can take place from one water molecule to another without difficulty. There is consequently a transfer of electricity which is independent of the normal speed of the hydrogen ion. In this manner it is possible to account for the unusually large conductance of the hydrogen ion. A similar movement of a proton can occur, at least to some extent, in other hydroxylic solvents, e.g., alcohols, and the results show that the hydrogen ion conductance in methanol and ethanol is appreciably larger than the values for other ions. In acetone the Grotthuss type of conduction is impossible, and the results in Table 152 show that the hydrogen ion in acetone, where it is probably $(\text{CH}_3)_2\text{C}=\text{OH}^+$, has a value of $\lambda\eta_0$ similar to that of other ions. The high conductance of the hydroxyl ion in water is probably also to be attributed to chain conduction of the Grotthuss type.⁶

Applications of Ion Conductances.—The equivalent conductance of a weak electrolyte cannot be obtained from direct conductance measure-

ments (p. 894), but it may be derived from the known conductances of the constituent ions. The value of Λ_0 for acetic acid, for example, is the sum of the ion conductances of H^+ and CH_3COO^- , the latter being obtained from measurements on sodium acetate or other highly ionized salt of acetic acid; hence from Table 151,

$$\Lambda_0(CH_3COOH) = \lambda_{H^+} + \lambda_{C_2H_3O_2^-} = 349.8 + 40.9 = 390.7 \text{ ohms}^{-1} \text{ cm.}^2 \text{ at } 25^\circ.$$

An alternative method, based on the law of independent migration of ions, is to utilize the relationship

$$\Lambda_0(CH_3COOH) = \Lambda_0(HCl) + \Lambda_0(CH_3COONa) - \Lambda_0(NaCl).$$

The three terms on the right hand side are equivalent to $\lambda_{H^+} + \lambda_{Cl^-}$, $\lambda_{Na^+} + \lambda_{CH_3COO^-}$, and $\lambda_{Na^+} + \lambda_{Cl^-}$, respectively, which together give $\lambda_{H^+} + \lambda_{CH_3COO^-}$, as required for acetic acid. The equivalent conductance of the latter at infinite dilution can thus be evaluated from the data for three strong electrolytes.

The value of Λ_0 for a sparingly soluble salt is usually determined from ion conductances; the factor $\frac{1}{2}$ in the case of barium sulfate is used to indicate that 1 g. equiv., and not 1 mole, is concerned.

$$\Lambda_0(\frac{1}{2}BaSO_4) = \lambda_{\frac{1}{2}Ba^{++}} + \lambda_{\frac{1}{2}SO_4^{--}} = 63.6 + 78.8 = 143.4 \text{ ohms}^{-1} \text{ cm.}^2 \text{ at } 25^\circ,$$

and

$$\Lambda_0(AgCl) = \lambda_{Ag^+} + \lambda_{Cl^-} = 61.92 + 76.34 = 138.3 \text{ ohms}^{-1} \text{ cm.}^2 \text{ at } 25^\circ.$$

Determination of Solubilities.—If s is the solubility, in g. equiv. per liter, of a sparingly soluble salt and the specific conductance of the saturated solution is κ , then by (7) its equivalent conductance is $1000 \kappa/s$. The solution is so dilute that this may be assumed equal to Λ_0 , and since the latter can be obtained as the sum of the ion conductances, it is apparent that the solubility of the salt can be estimated from a knowledge of κ ; thus,

$$s = 1000 \kappa / \Lambda_0. \quad (14)$$

The specific conductance of a saturated solution of silver chloride in water is $3.41 \times 10^{-6} \text{ ohm}^{-1} \text{ cm.}^{-1}$ at 25° C. , and if 1.60×10^{-6} is deducted for the conductance of the water, the value of κ is $1.81 \times 10^{-6} \text{ ohm}^{-1} \text{ cm.}^{-1}$. It was seen above that for silver chloride Λ_0 is $138.3 \text{ ohms}^{-1} \text{ cm.}^2$; hence,

$$s = 1000 \times 1.81 \times 10^{-6} / 138.3 = 1.31 \times 10^{-5} \text{ g. equiv. per liter at } 25^\circ.$$

The solubilities of other sparingly soluble salts and oxides have been determined in the same manner, but the results are not always reliable because of uncertainties as to the nature of the ions formed and on account of hydrolysis which may be appreciable in such dilute solutions (cf. p. 986).

Conductometric Titration.—If an alkali is added to a strong acid the highly conducting hydrogen ions of the latter are replaced by metallic ions having a much lower conductance; the conductance of the solution will therefore fall. When neutralization is complete the further addition of alkali will cause the conductance to increase, and so the neutral point is marked by a

minimum in the conductance. If a moderately weak acid, e.g., acetic acid, is neutralized by a strong base then, owing to the effect of the common ion in repressing dissociation, the conductance of the solution first falls, but it soon commences to increase because the conducting power of the highly ionized salt exceeds that of the weak acid which it replaces. Immediately the equivalence point is passed there is a more rapid increase of conductance because of the contribution of the hydroxyl ions from the excess alkali. If the acid is very weak, e.g., phenol, the conductance of the original solution is very small and it rises from the beginning of neutralization. The curves representing the variation of conductance of the acid with the amount of alkali added for the three cases, viz., (I) strong acid, (II) moderately weak acid, and (III) very weak acid, are shown

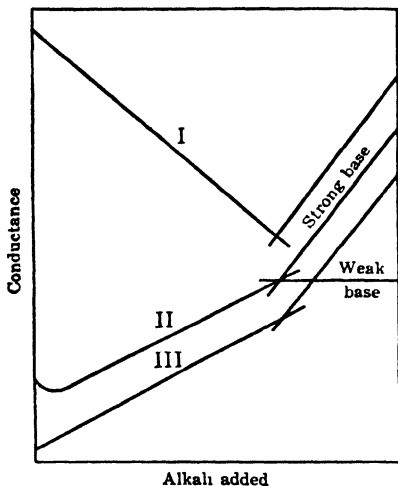
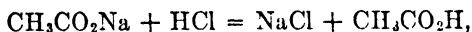


FIG. 232. Conductometric titration

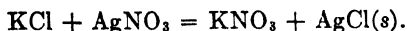
in Fig. 232. It is evident from these curves that conductance measurements may be utilized to determine the end-points in acid-base titrations. Actual conductances are not necessary, for only relative values are required, and these may be obtained directly from the Wheatstone bridge readings or in other ways. Only six or eight readings need be taken and two lines drawn through them to give the position of intersection. If there is little volume change during the titration, the alkali being about ten times as concentrated as the acid, the lines are practically straight. When titrating weak acids the sharpness of the end-point may be improved by using a weak base; this has very little conductance and so the plot is almost horizontal after the equivalence point has been passed (Fig. 232).

Conductometric methods can be used for the analysis of very dilute or colored solutions where ordinary indicators give poor results, and mixtures of strong and weak acids can be readily estimated.

When a strong acid, e.g., hydrochloric acid, is added to the salt of a weak acid, e.g., sodium acetate, thus



there is little change in conductance, because the acetic acid contributes very little and the sodium chloride has much the same conductivity as the acetate which it replaces. After the end-point of the reaction, however, the conductance rises because of the excess of strong acid. Similarly, there is an almost constant conductance followed by an increase in a precipitation reaction, e.g.,



If both the products are sparingly soluble, as in the titration of many sulfates by baryta solution, e.g.,



the conductance falls until the sulfate is completely precipitated and then increases because of the free barium hydroxide. It is evident that in all these reactions the end-point can be determined for analytical purposes by conductometric methods.

Determination of Basicity, etc.—W. Ostwald (1887) discovered an empirical relationship between the basicity b of an acid and the equivalent conductances of its sodium or potassium salt at dilutions of 1024 liters (Λ_{1024}) and 32 liters (Λ_{32}) per g. equiv.; thus,

$$\Lambda_{1024} - \Lambda_{32} = 10.8b. \quad (15)$$

The formula has proved particularly applicable to organic acids; it is not so satisfactory for inorganic acids, although it has been used to show that persulfuric acid is dibasic whereas permanganic acid is monobasic.

Most ions, with the exception of those having long chains and some complex ions, have ion conductances of about $60 \text{ ohms}^{-1} \text{ cm.}^2$ at 25°C. , and this fact may be applied to throw light on the mode of ionization of electrolytes. It has been found of particular value in connection with Werner coordination compounds, to determine whether a halogen atom, or other negative group, is attached in a covalent or electrovalent manner. The molar conductance of a uni-univalent salt MA will be about $120 \text{ ohms}^{-1} \text{ cm.}^2$, but that of a uni-bi-, or bi-uni-valent electrolyte, M_2A or MA_2 , will be about $240 \text{ ohms}^{-1} \text{ cm.}^2$. Consider M_2A , for example; this gives two M^+ ions which contribute approximately 120, and the A^{--} ion contributes the same amount, since the equivalent conductance of $\frac{1}{2}\text{A}^{--}$ is about $60 \text{ ohms}^{-1} \text{ cm.}^2$. In the same manner it can be shown that salts of the type M_3A or MA_3 will have molar conductances of the order of 360, whereas for M_4A or MA_4 they will be about $480 \text{ ohms}^{-1} \text{ cm.}^2$ at 25°C. The solutions must be relatively dilute, e.g., 1 mole in 1000 liters, so that dissociation is almost complete. As an illustration, reference may be made to the compound $\text{Pt}(\text{NH}_3)_6\text{Cl}_4$; the molar conductance is about $520 \text{ ohms}^{-1} \text{ cm.}^2$ at a dilution of 1024 liters, so that it is evidently a salt of the MA_4 type, and ionization must occur to yield $\text{Pt}(\text{NH}_3)_6^{++++}$ and 4Cl^- ions.

The Conductance Ratio.—When all the ions that can be formed from 1 g. equiv. of an electrolyte take part in conducting the current, the equivalent conductance is then Λ_0 which, by (11), is equal to $k(u_+ + u_-)$; this condition is reached only at extreme dilutions. In the development of the theory of ionic dissociation (p. 887) Arrhenius supposed that at appreciable concentrations a fraction α only of the 1 g. equiv. is dissociated into ions, and if the velocities of the ions are constant the equivalent conductance Λ of the particular solution will be $\alpha k(u_+ + u_-)$; hence,

$$\frac{\Lambda}{\Lambda_0} = \frac{\alpha k(u_+ + u_-)}{k(u_+ + u_-)} = \alpha, \quad (16)$$

and so the degree of dissociation might be calculated from conductance measurements. The fundamental error in this deduction is the fact that ionic velocities are *not* constant, because of the mutual interaction of the ions at appreciable concentrations (p. 903). According to modern views α as given by (16) has no connection with the degree of dissociation except for weak electrolytes; in fact it probably gives a better measure

of the change in the speeds of the ions than of their number. For this reason the quantity Λ/Λ_0 is no longer called the degree of dissociation or ionization, but the noncommittal name **conductance ratio** is employed. In weak electrolytes the number of ions is relatively small and so also is the interionic action which reduces the ionic speeds; for such solutions the velocities do not change very much with concentration and the conductance ratio is a reasonably good measure of the degree of dissociation or ionization. To obtain more accurate results an allowance can be readily made for the small change in the ionic speeds. If u and v are the velocities of the ions in the solution of appreciable concentration, and u_0 and v_0 are the values at infinite dilution, then (16) should be written

$$\frac{\Lambda}{\Lambda_0} = \frac{\alpha k(u + v)}{k(u_0 + v_0)}; \quad (17)$$

$$\therefore \alpha = \frac{\Lambda k(u_0 + v_0)}{\Lambda_0 k(u + v)}. \quad (18)$$

Now, $k(u_0 + v_0)$ is equal to Λ_0 , and $k(u + v)$ is the sum of the conductances of the two ions in solutions at the same ionic concentration as the solution under consideration; representing this by Λ' , (18) becomes

$$\alpha = \Lambda/\Lambda'. \quad (19)$$

The quantity Λ' may be obtained by the use of theoretical equations (p. 907), or it may be derived directly from experimental data, as the following illustration will show. Consider phosphoric acid, ionizing as a monobasic acid into H^+ and $H_2PO_4^-$ ions; then, at $18^\circ C.$,

$$\begin{aligned} \Lambda_0(H_3PO_4) &= \Lambda_0(HCl) + \Lambda_0(NaH_2PO_4) - \Lambda_0(NaCl) \\ &= 378.3 + 70.0 - 109.0 = 339.3. \end{aligned}$$

The measured value of Λ for phosphoric acid at a concentration of 0.1 g. equiv. per liter is 96.5, and so by (16) the approximate degree of dissociation is $96.5/339.3$, i.e., 0.285, and the concentration of the ions is about 0.0285 g. equiv. per liter. It is now necessary to find the sum of the conductances of H^+ and $H_2PO_4^-$ ions at that ionic concentration; this is obtained from data on the conductances of HCl , NaH_2PO_4 , and $NaCl$. These are strong electrolytes and may be assumed to be completely ionized at a concentration of 0.0285 g. equiv. per liter, so that this is the concentration at which the conductances are required. The measured equivalent conductances are 363.4, 59.3 and 98.6, respectively, so that

$$\Lambda'_{(H_3PO_4)} = 363.4 + 59.3 - 98.6 = 324.1.$$

The more accurate value of the degree of dissociation of phosphoric acid at 0.1 g. equiv. per liter is thus, by (19), $96.5/324.1$, i.e., 0.298. The result may be refined still further by calculating a more accurate value of Λ' from measurements at concentrations of 0.0298, instead of the former figure 0.0285; the correction is very small, however, and may be

ignored. In this particular instance the difference between the approximate (0.285) and more exact (0.298) results for the degree of dissociation is relatively small; with more concentrated solutions, or stronger electrolytes, the discrepancy is much greater. It appears, however, that as a first approximation Λ/Λ_0 may be used to evaluate the degree of dissociation of a weak electrolyte, provided the solution is not too dilute or too concentrated.⁸

The Interionic Attraction Theory.—For many years it had been realized that a weakness of the Arrhenius method of calculating the degree of dissociation of a strong electrolyte lay in the failure to take into consideration interionic forces (J. J. van Laar, 1895; A. A. Noyes, 1904; W. Sutherland, 1905; N. Bjerrum, 1909; S. R. Milner, 1912), and the tentative suggestion had been made that such electrolytes were completely dissociated at all dilutions, the falling off in conductance with increasing concentration being due mainly to the decreased velocity of the ions. The development of the study of crystals by X-rays and of the electronic theory of valence, however, provided a satisfactory basis for this view. It is now generally accepted that even in the solid state a salt consists of positive and negative ions held together by electrostatic forces, and that a nonionized molecule, e.g., of sodium chloride, does not exist, at least in the crystal. It is reasonable to assume, therefore, that when a salt is dissolved in water, and the ions become free to move and conduct electricity, it is still *completely ionized*. The electrostatic attraction existing in the crystal persists in the solution, although it is opposed by other factors, with the result that at appreciable concentrations there may be a definite pairing of ions, or a continual interchange of partners between ion-pairs of this kind. The consequence of this is that although ionization may be complete, *dissociation will be incomplete* except in very dilute solutions. Under the latter conditions the ions are comparatively far apart and the interionic attractions are negligible; the solution then behaves as if the electrolyte were completely dissociated, as well as completely ionized. Nevertheless, even in concentrated solutions the proportion of "undissociated" molecules in a strong electrolyte is very much less than was assumed by Arrhenius.

Quantitative Treatment of Theory of Electrolytes.—An attempt to treat the interionic attraction theory of electrolytes quantitatively was made by J. C. Ghosh (1918), with relatively little success, but the work of P. Debye and E. Hückel (1923) marked the commencement of a new era in electrochemistry. The fundamental idea underlying their deductions is that as a consequence of electrical attractions between positive and negative ions there are, on the average, more ions of unlike than of like sign in the neighborhood of any ion. Every ion may, therefore, be regarded as being surrounded by an **ionic atmosphere** of opposite charge. As long as the system is "stationary," that is to say, it is not exposed to an applied electric field or to a shearing force which tends to cause the

liquid to flow, the ionic atmosphere has central symmetry. When a current is passed through the solution, however, so that the ions of a particular sign move, say, to the right, then each ion during its motion will constantly have to build up its ionic atmosphere to the right, whereas the charge density to the left will die out. If this formation and destruction of the ionic atmosphere occurred instantaneously, there would be no net force of attraction on the moving ion, but in effect there is a definite **time of relaxation** during which the atmosphere to the right is building up to its equilibrium value and that to the left is decaying. Since the charge of the ionic atmosphere is opposite to that of the moving ion, there will be an excess of ions of opposite sign to the left and these will retard the motion. The influence on the velocity of the ion is known as the **relaxation effect**, or sometimes as the **asymmetry effect**, because it arises from the lack of symmetry in the electrical atmosphere of a moving ion. In addition to the foregoing, another factor will help to oppose the motion of the ions: this is the tendency of the applied E.M.F. to move the ionic atmosphere, with its associated molecules of solvent, in a direction *opposite* to that in which the ion, accompanied by solvent molecules, is itself moving. The additional retardation arising in this manner is called the **electrophoretic effect**, since it is analogous to that opposing the movement of a colloidal particle in an electric field (see p. 1238). Finally, the migration of the ion is opposed by the normal frictional resistance of the medium; this is determined by Stokes's law, and depends on the speed of the ion, its radius and the viscosity of the medium. When the ion moves with a steady velocity the three retarding forces are just balanced by the electrical force due to the applied E.M.F. At infinite dilution the asymmetry and electrophoretic effects are virtually zero, and the speeds of the ions, and hence the equivalent conductance, are determined only by the frictional force of the medium. The difference between the conductance at infinite dilution and at an appreciable concentration is, therefore, a direct consequence of the two electrical forces, assuming the viscosity of the solution to be little different from that at infinite dilution, i.e., that of the solvent.

On the basis of these arguments Debye and Hückel were able to derive an expression relating the observed equivalent conductance at a particular concentration to that at infinite dilution. The calculations were subsequently improved by L. Onsager (1926-27), and the result obtained, known as the Debye-Hückel-Onsager equation or, in brief, as the **Onsager equation**, may be written as

$$\Lambda = \Lambda_0 - \left(\frac{\epsilon \kappa}{300} \cdot \frac{z_+ + z_-}{6\pi\eta} F + \frac{\epsilon^2 \kappa}{6DkT} \Lambda_0 w \right), \quad (20)$$

where ϵ is the electronic charge, κ is the effective radius of the ionic atmosphere, the value of which is given by (119), p. 958, z_+ and z_- are the (numerical) valences of the ions, η is the viscosity of the medium, D

is its dielectric constant, k is the Boltzmann constant, i.e., the gas constant per molecule, while F and T are, as usual, the faraday and the absolute temperature, respectively. The factor w is defined by

$$w = 2qz_+z_-/(1 + \sqrt{q}), \quad (21)$$

where

$$q = z_+z_-(\lambda_+ + \lambda_-)/(z_+ + z_-)(z_+\lambda_- + z_-\lambda_+), \quad (22)$$

λ_+ and λ_- being, as before, the equivalent conductances of the indicated ions. The first term in the parentheses in (20) represents the decrease in conductance due to the electrophoretic effect, and the second term results from the relaxation or asymmetry effect. The former is seen to be dependent on the viscosity of the solvent, while the latter is influenced by its dielectric constant.⁹

It will be seen later that the effective radius κ of the ionic atmosphere is related to the square root of the concentration of the ions and the dielectric constant of the medium, apart from universal constants. If the latter, as well as those in (20), are inserted, it is found that

$$\Lambda = \Lambda_0 - \left[\frac{29.15(z_+ + z_-)}{(DT)^{1/2}\eta} + \frac{9.90 \times 10^5}{(DT)^{1/2}} \Lambda_0 w \right] \sqrt{c(z_+ + z_-)}, \quad (23)$$

where c is concentration of the solution in g. equiv. per liter, *assuming the electrolyte to be completely in the form of free ions*. In the simple case of a uni-univalent electrolyte, z_+ and z_- are unity, and w is $2 - \sqrt{2}$; (23) then reduces to

$$\Lambda = \Lambda_0 - \left[\frac{82.4}{(DT)^{1/2}\eta} + \frac{8.20 \times 10^5}{(DT)^{1/2}} \Lambda_0 \right] \sqrt{c}, \quad (24)$$

where c is used in place of c , since concentrations in moles and g. equiv. are now identical. The general form of (24) may be represented by

$$\Lambda = \Lambda_0 - (A + B\Lambda_0) \sqrt{c} \quad \text{or} \quad \Lambda_0 - \Lambda = (A + B\Lambda_0) \sqrt{c}, \quad (25)$$

where A and B are constants dependent only on the temperature and the solvent. The calculated values at 25° for water, and methyl and ethyl alcohols are as follows:

	A	B
Water	60.2	0.229
Methyl alcohol	156.1	0.923
Ethyl alcohol	89.7	1.33

Verification of the Onsager Equation.—Since the form of (25) requires the equivalent conductance to vary in a linear manner with the square root of the concentration, it is in general agreement with experiment for dilute solutions in solvents of dielectric constant greater than about 20. The presence of the square root term is merely the result of the application of Coulomb's law to evaluate the interionic forces involved in the ionic atmosphere, and so the fact that (23) and (25) are of the correct form

provides little support for the main theory. The real test lies in a comparison of the slope of the line giving the variation of Λ with \sqrt{c} with that calculated from the values of A and B . In water and methyl alcohol a number of simple salts, especially alkali halides, give slopes in excellent agreement with those derived from (25), and even where differences exist they are not of large magnitude, especially in dilute solutions. In Fig. 233, for example, are plotted the measured equivalent conductances of a number of strong electrolytes in water at 25° C.; the broken lines show

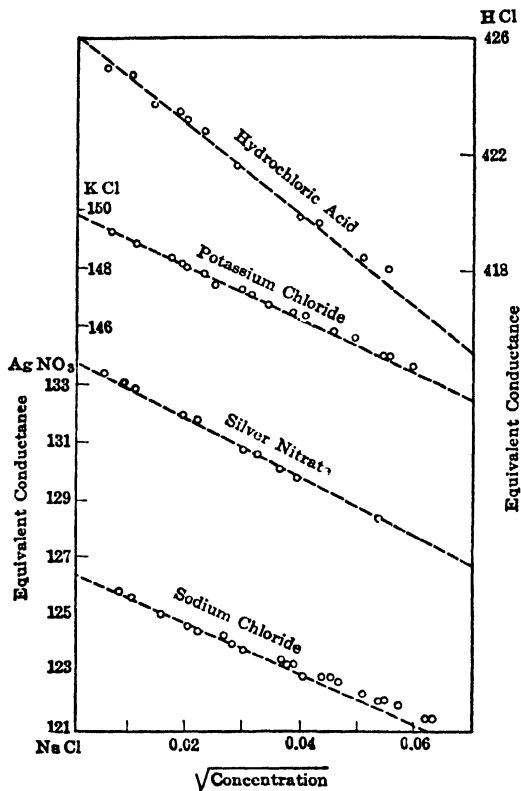


FIG. 233. Test of Onsager equation

the theoretical slopes according to the Onsager equation. Fairly good agreement has also been obtained for some salts dissolved in ethyl alcohol, and in nonhydroxylic solvents, e.g., nitriles, nitro-compounds and ketones, but for other salts large deviations have been observed. The fact that there are many instances in which (25) represents almost exactly the variation of conductance with concentration in dilute solution, provides very powerful support for the general theory of interionic attraction,

based on the supposition of complete ionization *and* complete dissociation of strong electrolytes. Certain salts, e.g., sodium and potassium chlorides, give results in excellent agreement with the Onsager equation at low concentrations, but at appreciable concentrations the measured conductances are somewhat larger than those calculated. It is probable that these deviations are due to the neglect of certain terms, e.g., size of the ions and interaction between the various retarding effects, in the Debye-Hückel-Onsager treatment, and so do not constitute an obstacle to the theory.¹⁰

Incomplete Dissociation.—A much more serious problem is presented by many electrolytes which have conductances varying in a linear manner with \sqrt{c} , but the slopes of the lines are appreciably greater than required by (25); this means that the observed conductance is always lower than the calculated value. Weak electrolytes, e.g., weak acids and bases, are undoubtedly only partially ionized at all reasonable concentrations, and such solutions have conductances considerably lower than those given by the Onsager equation. It appears certain, therefore, that the smaller differences frequently encountered, particularly with salts of the transition metals, are to be explained by the fact that dissociation is incomplete and not all the ions possible are available for carrying the current. Some may be held together by valence forces in the form of nonionized molecules, such as exist in solutions of many acids and bases, but some will be present in the form of associated ion-pairs as a consequence of electrostatic attraction. This view has been developed as a means of calculating the degree of dissociation of an electrolyte, that is, the average fraction of the solute free to carry current, although it is impossible to distinguish between true nonionized molecules and associated ion-pairs.

In (25) the quantity Λ represents the conductance of 1 g. equiv. of electrolyte at concentration c , when *completely in the form of free ions*, the term $(A + B\Lambda_0)\sqrt{c}$ being the diminution from the value at infinite dilution due to interionic forces. If the fraction α only of the 1 g. equiv. is present as free ions, where α is the degree of dissociation in the general sense referred to above, then the concentration of free ions is αc , and the interionic term becomes $(A + B\Lambda_0)\sqrt{\alpha c}$.* The conductance of 1 g. equiv. of *free ions* at this concentration, which is the quantity designated by Λ' on p. 902, is then given by

$$\Lambda' = \Lambda_0 - (A + B\Lambda_0)\sqrt{\alpha c}. \quad (26)$$

Since 1 g. equiv. of electrolyte contains only α g. equiv. as free ions, the measured equivalent conductance Λ is $\alpha\Lambda'$, and α is equal to Λ/Λ' , as previously shown. To evaluate α and Λ' , (26) is written in the form

$$\Lambda' = \Lambda_0 - (A + B\Lambda_0)\sqrt{c\Lambda/\Lambda'}, \quad (27)$$

* It is important to note that the calculations leading to (25) do not necessarily imply complete ionization; the result should give the interionic effect in *any* electrolyte provided the concentration of *free ions* is used.

and provisional values of Λ' are obtained for different concentrations; the quantity Λ' under the square root sign is replaced by Λ_0 , an adequate value being obtained either by extrapolation or from the ion conductances. The approximate result for Λ' is then used in the $\sqrt{c\Lambda/\Lambda'}$ term and a more accurate value of Λ' computed; the process is continued until Λ' remains constant, and from this α at the given concentration may be determined (C. W. Davies, 1927).¹¹

Conductance at High Frequencies and High Voltages.—If a high frequency alternating current is applied to an electrolyte and the time of oscillation is equal to, or less than, the relaxation time of the ionic atmosphere (p. 904), the unsymmetrical charge distribution usually associated with an ion in motion cannot form completely. The retardation of the ionic velocity due to the asymmetry effect is thus reduced or annulled, and the conductance of the solution should be greater than the normal value. The time of relaxation is found, by calculation, to be of the order of $10^{-10}/c$ sec., and so for a 0.001 N solution it is about 10^{-7} sec., and the effect on the conductance of oscillations with a wave length of 1 meter, frequency 3.3×10^{-8} sec., should be noticeable. The existence of the high frequency effect was deduced theoretically by P. Debye and H. Falkenhagen (1928), and verified experimentally by H. Sack (1928) and others.

An increase in the equivalent conductance of an electrolyte was observed by M. Wien (1927) when using very high voltages, e.g., 20,000 volts per cm.; under these conditions an ion moves at a speed of the order of 10 cm. per sec., and so travels many times the thickness of the ionic atmosphere, which is about 10^{-8} cm. (p. 958), during the relaxation time. As a result the moving ion is virtually free from an oppositely charged atmosphere, and so neither the asymmetry nor the cataphoretic effect is operative; the equivalent conductance is, therefore, increased and approaches the value for infinite dilution. The theory of the influence of high voltages on conductance, called the **Wien effect**, has been worked out on the basis of the concept of interionic attraction, and the equations derived are in harmony with the experimental results. It is important to emphasize that the influence of high frequency currents and high voltages provides strong arguments in favor of the view that under normal conditions ionic velocities, and hence conductances, are below those for infinite dilution because of the effect of the ionic atmosphere which increases with increasing concentration.¹²

Absolute Velocities of Ions.—Suppose a dilute solution of concentration c g. equiv. per liter, i.e., $c/1000$ g. equiv. per cc., is placed in a cube of 1 cm. side with electrodes of 1 cm. squares at opposite faces and an E.M.F. of 1 volt applied; the conductance of the 1 cm. cube is the specific conductance (κ) by definition, and by (7), Λ is $1000 \kappa/c$. Since the solution is dilute it may be assumed that Λ is almost identical with Λ_0 and hence, by (10), with $\lambda_+ + \lambda_-$, so that

$$1000\kappa/c = \lambda_+ + \lambda_- \quad \text{and} \quad \kappa = c(\lambda_+ + \lambda_-)/1000. \quad (28)$$

By Ohm's law the current flowing is equal to the product of the E.M.F. and the conductance, which in this case is $1 \times \kappa$, i.e., $c(\lambda_+ + \lambda_-)/1000$

amp., and so during every sec. this number of coulombs flows through the 1 cm. cube. If u_+ and u_- are the actual ionic velocities, in cm. per sec., under a fall of potential of 1 volt per cm., then all the cations within a length u_+ cm. will pass across a given plane in one direction in 1 sec., while all the anions within a length u_- cm. will pass in the opposite direction. Suppose the plane has an area of 1 sq. cm.; all the cations in volume u_+ cc. and all the anions in u_- cc. will pass in opposite directions per sec.; that is a total of $(u_+ + u_-)c/1000$ g. equiv. of cations and anions will be transported in 1 sec. Since each g. equiv. of any ion carries 96,500 coulombs, that is one faraday (F), the total quantity of electricity carried per sec. will be $F(u_+ + u_-)c/1000$ coulombs. It has been found above that the quantity of electricity flowing is $c(\lambda_+ + \lambda_-)/1000$ coulombs per sec., and so

$$F(u_+ + u_-)c/1000 = c(\lambda_+ + \lambda_-)/1000; \quad (29)$$

$$\therefore F(u_+ + u_-) = \lambda_+ + \lambda_- \quad (30)$$

By (12), λ_+ is equal to ku_+ and λ_- to ku_- , and so k is equal to F , and hence

$$u_+ = \lambda_+/F \quad \text{and} \quad u_- = \lambda_-/F. \quad (31)$$

The absolute velocity of any ion under a fall of potential of 1 volt per cm., generally called the **ionic mobility**, is thus obtained, in cm. per sec., when the ion conductance, in ohms⁻¹ cm.², is divided by the faraday, i.e., 96,500 coulombs. The values obtained are for infinite dilution, and for appreciable concentrations they are probably somewhat less. A few typical ionic mobilities, derived from the experimental ionic conductances at 25° c., are given in Table 153; like the ionic conductances themselves,

TABLE 153. CALCULATED IONIC MOBILITIES AT 25°.

Cation	Cm./sec.	Anion	Cm./sec.
Hydrogen	36.2×10^{-4}	Hydroxyl	20.5×10^{-4}
Potassium	7.61	Sulfate	8.27
Barium	6.60	Chloride	7.91
Sodium	5.19	Nitrate	7.40

the mobilities increase by about 2 per cent for every degree rise of temperature.

Experimental Determination of Ionic Speeds.—In the first attempts to measure ionic velocities O. Lodge (1886) used a tube about 40 cm. long joining the anode and cathode compartments; this contained a conducting gel together with a material for indicating the movement of the ion being studied. For example, in determining the speed of the barium and chloride ions, using barium chloride in the anode and cathode vessels, the gel contained a small amount of silver sulfate. On the passage of current, barium and chloride ions traveled into the gel, from opposite directions, and their movement was followed by the visible precipitates of barium sulfate and silver chloride, respectively. In measuring the speed of hydrogen ions the gel contained sodium chloride as conductor, and phenolphthalein turned just pink by alkali; as the hydrogen

ion proceeded in the direction of the current it decolorized the phenolphthalein. The result obtained for this ion was in fair agreement with that given in Table 153, but for barium, strontium, chloride and bromide ions the observed velocities were one-third to one-fifth of those calculated from the ionic conductances by (31). According to W. C. D. Whetham (1893) the errors were due to a nonuniform potential gradient in the conducting gel, and to the failure to obtain sharp boundaries. In Whetham's experiments two solutions were used, the noncommon ion in one of them being colored; the specific conductances were almost the same so that the potential gradient was uniform. A sharp boundary between the two liquids was made in a relatively narrow vertical tube, and was maintained thus by having the denser solution below the other, and by arranging that the boundary moved in the direction of the solution containing the faster noncommon ion (cf. p. 914). The rate of movement of the boundary between colored and colorless solutions was measured, and from this the velocity of the colored ion was calculated for a potential gradient of 1 volt per cm. The same method was later used by Whetham (1895) for colorless ions, and the results were always of the same order as those calculated.¹³

Transference (or Transport) Numbers.—The quantity of positive electricity carried by the cations in one direction is proportional to their speed u_+ and their equivalent concentration c_+ ; similarly, the quantity of negative electricity moving in the opposite direction is proportional to c_-u_- . The total current carried, as seen on p. 909, is proportional to $c_+u_+ + c_-u_-$; hence the fraction of the total current carried by each of the ionic species, called the **transference (or transport) number**, t_+ or t_- , of that ion, is given by

$$t_+ = \frac{c_+u_+}{c_+u_+ + c_-u_-} \quad \text{and} \quad t_- = \frac{c_-u_-}{c_+u_+ + c_-u_-}, \quad (32)$$

for a solution containing a single electrolyte. In the simple case of a molecule of the latter forming only two ions, e.g., KCl, MgSO₄, etc., c_+ is equal to c_- , and hence

$$t_+ = \frac{u_+}{u_+ + u_-} \quad \text{and} \quad t_- = \frac{u_-}{u_+ + u_-}. \quad (33)$$

For the sake of simplicity, an electrolyte of this type will be assumed in most of the subsequent treatment.

As a result of the difference in the velocities of the two ions of an electrolyte different concentration changes occur in the neighborhood of the cathode and anode, respectively, and these changes may be utilized, as first shown by W. Hittorf (1853), before the introduction of the theory of electrolytic dissociation, to evaluate transference numbers. Consider an electrolyte MA, giving M⁺ and A⁻ ions with transference numbers t_+ and t_- , respectively; suppose current passes through the solution contained in a cell divided into three compartments, one surrounding the cathode, another the anode, and a third, or middle, compartment in which no concentration change occurs. If one faraday of electricity is passed, t_+

faradays are carried by t_+ g. equiv. of cations in one direction and t_- faradays in the other by t_- g. equiv. of anions; this is accompanied by the following changes in the three compartments.

Cathode Compartment (I)	Middle Compartment	Anode Compartment (II)
1 g. equiv. of M^+ is discharged.	t_+ g. equiv. of M^+ migrate to I	1 g. equiv. of A^- is discharged.
t_+ g. equiv. of M^+ migrate in.	t_- g. equiv. of A^- migrate from I	t_- g. equiv. of A^- migrate in.
t_- g. equiv. of A^- migrate out.	t_+ g. equiv. of M^+ migrate from II	t_+ g. equiv. of M^+ migrate out.
	t_- g. equiv. of A^- migrate to II	
Net Result:		
Loss of $1 - t_+ = t_-$ g. equiv. of M^+ .	No change of concentration.	Loss of $1 - t_- = t_+$ g. equiv. of A^- .
Loss of t_- g. equiv. of A^- . i.e., Net loss of t_- g. equiv. of MA.		Loss of t_+ g. equiv. of M^+ . i.e., Net loss of t_+ g. equiv. of MA.

It is seen, therefore, that

$$\frac{\text{No. of g. equiv. of electrolyte lost from anode compartment}}{\text{No. of g. equiv. of electrolyte lost from cathode compartment}} = \frac{t_+}{t_-}.$$

Further,

$$\frac{\text{No. of g. equiv. of electrolyte lost from anode compartment}}{\text{No. of g. equiv. deposited on each electrode}} = \frac{t_+}{1} = t_+.$$

If a coulometer is included in the circuit, then by Faraday's laws the same number of g. equiv. of material, no matter what its nature, will be deposited as in the cell under consideration; hence,

$$\frac{\text{No. of g. equiv. of electrolyte lost from anode compartment}}{\text{No. of g. equiv. of metal deposited in coulometer}} = t_+ \quad (34)$$

and, similarly,

$$\frac{\text{No. of g. equiv. of electrolyte lost from cathode compartment}}{\text{No. of g. equiv. of metal deposited in coulometer}} = t_- \quad (35)$$

It is thus possible to evaluate the transference numbers of the ions of a given electrolyte from measurements of the fall of concentration of the anode and cathode compartments. In general, the number of g. equiv. lost is proportional to the transference number, and hence to the velocity, of the ion migrating *away from* that compartment. Since $t_+ + t_-$ is equal to unity, it is not necessary in practice, except for confirmatory purposes, to examine the concentrations in both compartments; if the transference number of one ion is determined, that of the other is immediately known.

In the above derivation it has been assumed that neither the electrodes nor the water are attacked by the ions liberated, so that when they are discharged they may be considered as being completely removed from the compartments. When anions are discharged, however, they are not always deposited on the anode in a form that effectively removes them from the system. For example, in the electrolysis of silver nitrate with silver electrodes, it is true that silver is deposited on the cathode in a solid form, but NO_3 radicals are not removed in an analogous manner at the anode. The silver anode will dissolve in an amount equivalent to the NO_3^- that would have been discharged according to the foregoing arguments. The concentration of the anode compartment is thus increased by an equivalent amount of silver nitrate, and an allowance must be made when calculating the transference number. This can be readily done, for the quantity of metal dissolved from the anode is exactly equivalent to that deposited at the cathode of the cell and also on the cathode of a coulometer in the circuit. When alkali chloride solutions are studied the anode is of finely divided silver and the cathode of silver covered with

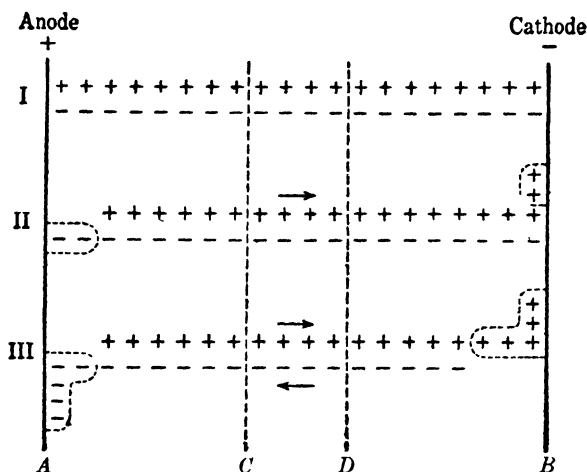


FIG. 234. Migration of ions

silver chloride. In this case the discharged chlorine at the anode combines with the silver to form the insoluble chloride, which is thus removed; at the cathode the alkali metal may be regarded as attacking the silver chloride to form alkali chloride which passes into solution. The correction must now be applied to the concentration of the cathode compartment.

Faraday's Law and Ionic Speeds.—It may be difficult to see how equivalent quantities of different ions are liberated at the two electrodes in a given solution, as required by Faraday's second law, in spite of the possible difference in the speeds of the ions moving towards the respective electrodes. A simple method

of proving that this must be so is by a consideration of the diagram in Fig. 234. Suppose a number of positive and negative ions are distributed evenly between two electrodes *A* and *B*, anode and cathode, respectively. The space between the electrodes is considered to be divided into the three sections mentioned above by two imaginary partitions, so that the section *AC* may be regarded as the anode compartment and *DB* as the cathode compartment. At the commencement of electrolysis the concentration of anions and cations is the same in both these sections, as shown in Fig. 234, I. Suppose that the cations *only* are able to move, and that under the influence of an applied E.M.F. two of these ions move from left to right; the condition attained will then be as at Fig. 234, II. At each electrode there are two ions unpaired and these must be considered to be discharged; thus, although only the positive ions are able to move equivalent amounts of the two ions are discharged at the respective electrodes. This condition actually arises in certain solid and fused electrolytes, where all the current is carried by one ion. Suppose, further, that while two cations move in one direction, three anions move in the other, so that the ionic velocities are in the ratio of 2 to 3; the result is shown in Fig. 234, III, where it is seen that five ions are discharged at each electrode. There is thus no difficulty in correlating Faraday's law with the unequal speeds of the ions.

It will be observed, incidentally, that in case II the concentration of the anode compartment has been diminished by two equiv. of electrolyte whereas that of the cathode section is unchanged; in case III, however, the concentration of the cathode compartment has decreased by three equiv. From these and other instances it is apparent that the amount of electrolyte lost from each compartment is proportional to the speed of the ions migrating away from it. This conclusion is identical with that reached above, and in fact equations (31) to (35) can be derived, although less explicitly, from a consideration of the results in Fig. 234.

Moving Boundary Method.—A more direct method for studying transference numbers is based on the principle used by Lodge, and improved by Whetham, to determine the actual speeds of ions. If the velocities of both anion and cation can be measured with reasonable accuracy by observations of a moving boundary, it is evident that transference numbers can be evaluated directly. Advances in technique were made by D. O. Masson (1899), B. D. Steele (1901–06) and R. B. Denison (1906), but it is only in relatively recent times that the work of D. A. MacInnes (1923 *et seq.*) has made the procedure simple and precise. The principle of the moving boundary method may be explained with the aid of Fig. 235. If an electrolyte *MA*, e.g., potassium chloride, is to be studied its solution is placed in between those of two other salts *M'A* and *MA'*, e.g., lithium chloride

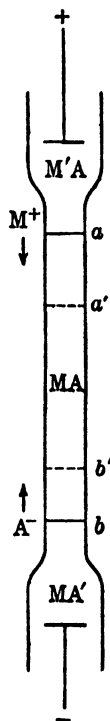


FIG. 235. Principle of the moving boundary method

and potassium acetate, known as indicator solutions, each having an ion in common with the first salt. If possible the three solutions are arranged in order of increasing density downwards to avoid mixing effects due to gravity. The anode is placed in the solution $M'A$ and the cathode in MA' , as shown. In order that the boundaries a and b , between $M'A$ and MA , and between MA and MA' , respectively, may remain distinct, the first requirement is that the speed of the ion M' should be less than that of M , and the speed of A' should be less than that of A . Under these conditions M' does not overtake M at a , and A' does not overtake A at b , and the boundaries do not become blurred. When an E.M.F. is applied a moves to a' while b moves to b' , and so the distances aa' and bb' are related to u_+ and u_- , the speeds of the ions M and A , respectively. Since the solution MA is homogeneous, its resistance is the same throughout and the fall of potential between a and b will be uniform. The ions M and A thus move under the same potential gradient, and the distances travelled must be directly proportional to their speeds; thus $aa'/bb' = u_+/u_-$,

$$\therefore t_+ = \frac{u_+}{u_+ + u_-} = \frac{aa'}{aa' + bb'} \quad \text{and} \quad t_- = \frac{bb'}{aa' + bb'}. \quad (36)$$

Another condition for the boundaries to remain sharp is that the cations M and M' should have the same speed near the junctions, and that the anions A and A' should also have equal velocities; it was shown by F. W. Kohlrausch (1897), and others, that this will hold good at any boundary if t/c , where c is the concentration, is the same for the two ions of the same sign. The adjustment of the indicator ion concentration so as to satisfy this requirement makes the potential gradients in the two solutions constituting the boundary such as to equalize the speeds of the M and M' ions at one boundary, and of A and A' at the other. It can be deduced theoretically that the Kohlrausch "regulating function" for a sharp boundary should be established automatically soon after the passage of current commences, and consequently the actual concentration of the indicator solutions should be immaterial. It is found in practice, however, that owing to the disturbing effects of convection, resulting from temperature and concentration gradients in the indicator solutions, the results are not independent of the concentration of the latter. It is advisable, therefore, to make the initial values as close as possible to those required by the Kohlrausch function. This can only be done after trial experiments to give provisional data concerning the transference numbers.

In the method of MacInnes only one boundary is observed, so that the necessity of finding two suitable indicator solutions is obviated; the method of calculation is as follows. If 1 faraday, i.e., 96,500 coulombs, of electricity flows through a solution, t_+ g. equiv. of cation must pass any given point in one direction, while t_- g. equiv. of anion pass in the opposite direction. Suppose ϕ is the volume of solution containing 1 g. equiv. of electrolyte; then during the passage of one faraday (F) the cation bound-

ary must sweep through a volume ϕt_+ . For the passage of Q coulombs, therefore, the same boundary will move through a volume $Q\phi t_+/F$. If a is the area of cross section of the tube in which the boundary is formed and l is the distance it moves, then the volume passed through is la , so that $la = Q\phi t_+/F$;

$$\therefore t_+ = laF/Q\phi. \quad (37)$$

In accurate work corrections must be applied for the changes in volume due to chemical reactions at the electrodes and to ionic migration.

Experimental Methods.—In his original work Hittorf used short, wide tubes separated by porous partitions, and measured the changes of concentration in the vicinity of the electrodes. Although the apparatus has been very considerably altered, the principle of determining transference numbers from concentration changes is still called the Hittorf method. A definite quantity of electricity, determined by a coulometer in the circuit, is passed through a

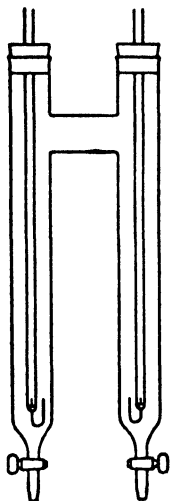


FIG. 236, A. Apparatus of Nernst

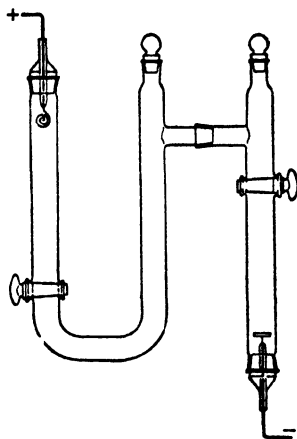


FIG. 236, B. Apparatus of Jones and Dole

solution of known concentration contained in a form of apparatus which permits of virtual, or real, separation into three compartments after electrolysis. The concentration in the anode or cathode compartment, or both, is then re-determined, and so also is that of the solution in the middle section. The latter should remain unaltered, but if it does not then it is evident that the concentration changes have extended too far as a result of either too prolonged passage of current, or of mixing of the electrolyte due to mechanical vibration, diffusion or convection effects. Precautions should be taken to avoid these sources of error.

Many forms of apparatus have been devised for the measurement of transference numbers by the Hittorf method. The apparatus depicted in Fig. 236, A is a modified form of one used by W. Nernst (1888), which is frequently em-

ployed for laboratory work, while that in Fig. 236, *b* is a schematic representation of the design employed by G. Jones and M. Dole (1929) for precision work on alkali halides. The general experimental method is as follows. The solution to be studied, whose *weight* concentration is known, is poured into the vessel and the anode and cathode fitted firmly. If the electrolyte is silver nitrate then silver electrodes may be used, and similarly for salts of other metals capable of being deposited on the cathode and dissolved at the anode, the metal itself can be used for the electrodes. With certain solutions, e.g., hydrochloric acid, sodium hydroxide and alkali halides, there is a possibility that gases will be liberated at one or both electrodes; the disturbance thus caused and the acid or alkali set free will vitiate the experiment. W. Hittorf (1858) used cadmium to avoid the liberation of chlorine at the anode, and cathodes of mercury covered with concentrated solutions of zinc chloride or copper nitrate have been employed to prevent the evolution of hydrogen. In the latter cases only the changes in concentration of the anode compartment can be utilized for the calculation of transference numbers. For the electrolysis of alkali halides the best electrodes are finely divided silver as anode, and silver covered with silver halide as cathode (p.912). The apparatus is now placed in a thermostat and the electrodes connected in series with a copper or silver voltameter; a current of 0.01 to 0.02 amp. is passed for two or three hours. At the conclusion of this time a sufficient quantity of solution, believed to contain all that has changed in concentration, is removed, weighed and analyzed. The adjacent solution from each side must prove on analysis to have the same concentration as the original electrolyte. The amount of metal deposited in the coulometer is also determined. In evaluating the loss of electrolyte from anode and cathode compartments allowance must be made for the quantity, if any, dissolved from anode or cathode. The loss is then calculated in terms of the weight of solute, before and after electrolysis, associated with the weight of water present in the solution run off at the end of the experiment. The transference numbers are then evaluated by means of (34) or (35), or both.

To facilitate the separation of anode and cathode compartments at the conclusion of electrolysis, ground-glass plugs and wide stopcocks have been employed (Fig. 236, *b*). Diffusion from one compartment to the other may be prevented by placing a wide U-tube between the limbs of the apparatus in Fig. 236, *a*; this U-tube is regarded as the middle compartment and the concentration of electrolyte in it should remain unchanged. Errors due to convection currents can be reduced by putting bends in the apparatus in suitable positions.¹⁴

The chief difficulty in the moving boundary method lies in the establishment of an initial sharp boundary. In the "sheared boundary" of D. A. MacInnes and T. B. Brighton (1925) two accurately fitting glass discs are connected to a vessel containing the experimental electrolyte and to a graduated tube containing the indicator solution, respectively. By a rotatory movement of the discs the two liquids are brought into contact; the protruding drops in each case are sheared off, and a sharp boundary is formed. The boundary, which can be a rising or a falling one according to convenience, moves in the graduated tube when current is flowing and the times are noted as it passes each 0.5 cm. graduation. The nature of the electrode material is determined by the same general considerations as are applicable to the Hittorf method. During the course of the experiment the current strength tends to decrease and continuous adjustment is necessary to keep it constant; this is desirable in order that the

boundary may move at a uniform rate. The tedious nature of this operation can be avoided by the use of an automatic current regulator. An important improvement in the moving boundary method is the introduction of the "autogenic boundary" (H. P. Cady and L. G. Longworth, 1929). No special indicator solution is required, but a metal is chosen to act as anode and to provide indicator ions when it dissolves under the influence of current. For example, with nitrate solutions a silver anode may be used and with chloride solutions one of cadmium; the silver nitrate or cadmium chloride, respectively, which is formed acts as the indicator. It is claimed that automatic adjustment of indicator concentration occurs, so that the Kohlrausch condition (p. 914) is satisfied, and no special care is necessary to produce a sharp boundary.

Transference Numbers: Results.—Provided the experiments are made carefully, identical results are obtained by the Hittorf and moving boundary methods, and the values are independent of the applied E.M.F. or current strength. Transference numbers do, however, vary with the concentration of the solution and with temperature; some results of cation transference numbers at 25° C., for various electrolytes at a series of concentrations, are quoted in Table 154. The variation of transference

TABLE 154. CATION TRANSFERENCE NUMBERS AT 25° C.
Concentration in G. Equiv. per Liter

Electrolyte	0.01	0.02	0.05	0.10	0.20	0.50
HCl	0.825	0.827	0.829	0.831	0.834	—
KCl	0.490	0.490	0.490	0.490	0.489	0.489
NaCl	0.392	0.390	0.388	0.385	0.382	—
AgNO ₃	0.465	0.465	0.466	0.468	—	—
BaCl ₂	0.440	0.437	0.432	0.425	0.416	0.399
CdI ₂ (18°)	0.444	0.442	0.396	0.296	0.127	0.003

number with concentration may be due to many causes: the influence of ionic hydration and of the formation of complex ions will be considered shortly, but in addition there are the changes in velocity of the ions as a result of interionic attraction. If the speeds of the anion and cation vary to relatively different extents as the concentration is altered, the transference numbers will also vary; this is probably the main cause of the changes in the values recorded above for electrolytes of the type of the alkali halides. According to G. Jones and M. Dole (1929) the empirical equation

$$t = \frac{A}{1 + B\sqrt{c}} - 1, \quad (38)$$

where A and B are constants, represents the variation of transference number with concentration for several electrolytes; the parameters A and B are chosen to fit the results, although A is not entirely arbitrary as it is equal to $t_0 + 1$, where t_0 is the transference number at infinite dilution, when c is zero. An equation of the form of (38) can be readily deduced from the Onsager equation (p. 905), but the theoretical values

of A and B differ from those found by experiment. The reason for this is that the transference measurements were made in solutions that were too concentrated for the Onsager equation to be applicable. For dilute solutions, that is, when c is small, (38) reduces to

$$t = t_0 - AB\sqrt{c}, \quad (39)$$

which is identical in form with the empirical relationship of G. Scatchard (1925). It appears from (39) that the plot of the transference number against the square root of the concentration should be a straight line for dilute solutions; this is in approximate agreement with experiment, and at very high dilutions the limiting slope approaches that required by the Onsager equation.

It was mentioned on p. 896 that the differences between ion conductances, and hence ionic velocities, decrease as the temperature is raised; this means that at sufficiently high temperatures the two ions of a given electrolyte would carry equal amounts of current, and the transference numbers would be 0.5. It is found in practice that as the temperature is raised transference numbers of all ions do in fact tend toward 0.5; those having smaller values, e.g., the sodium ion in sodium chloride, increase as the temperature is raised, whereas those with larger values, e.g., the hydrogen ion in hydrochloric acid, decrease. When the transference numbers of the ions are approximately 0.5, as in potassium chloride, the values are only slightly affected by temperature.

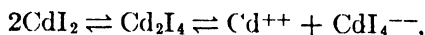
It is perhaps unnecessary to emphasize that the transference number, even at infinite dilution, differs from the ion conductance or ionic mobility in not being a constant for the given ion. Since the transference number is the *fraction* of the total current carried by the ion, it is clearly dependent both on its own velocity and on that of the other ion constituting the electrolyte. The connection between the transference number and the ion conductance may be derived by utilizing the relationships $\lambda_+ = ku_+$ and $\lambda_- = ku_-$, given on p. 895, whence it follows that

$$t_+ = \frac{u_+}{u_+ + u_-} = \frac{\lambda_+}{\lambda_+ + \lambda_-} = \frac{\lambda_+}{\Lambda_0}, \quad (40)$$

assuming the transference numbers, etc., to be the extrapolated values for infinite dilution; similarly, $t_- = \lambda_-/\Lambda_0$. It is thus possible to calculate limiting transference numbers from a knowledge of the conductances at infinite dilution of the ions concerned. In practice, the reverse procedure, namely the use of transference numbers to evaluate ion conductances, is actually employed. The transference number is extrapolated to zero concentration, and the product of this value and Λ_0 for the electrolyte gives the equivalent conductance of the ion at infinite dilution (Table 151).

Abnormal Transference Numbers.—Examination of Table 154 shows that the data for cadmium iodide are very abnormal; the transference number of the cadmium ion has fallen to 0.003 at 0.5 N, and at still higher

concentrations it is apparently negative. This anomalous observation was explained by W. Hittorf (1859) by postulating the existence of a complex anion, e.g., CdI_4^{--} , whose amount increases with the total concentration of cadmium iodide in solution. If ionization occurs in the following manner:



then cadmium in the form of Cd^{++} ions migrates out of the anode compartment while it is simultaneously replaced by CdI_4^{--} ions. If these two ions are present in similar amounts and have similar velocities the fall in concentration of the anode compartment may be negligible, so that the transference number of the cation will be apparently zero. In dilute solution the complex anion dissociates into simple ions and normal transference numbers are obtained.

Qualitative observations of ionic transference have been frequently employed in the study of complex ions; for example, in the electrolysis of potassium argentocyanide or ferrocyanide solutions the silver or iron, respectively, has been observed to migrate exclusively in the direction of the *anode*, showing that these elements are present almost entirely in the form of stable complex anions. In the ferrocyanide solution analysis shows that six CN radicals migrate with each atom of iron, so that the complex ion has the formula $[\text{Fe}(\text{CN})_6]_n$. The formula $[\text{Ag}(\text{NH}_3)_2]_n$ has been derived in an analogous manner for the argentammine ion. Transference measurements have proved the existence of complex anions containing magnesium, zinc or cadmium in solutions of the sulfates of these metals in the presence of alkali metal or ammonium sulfates (J. W. McBain, 1928-30).

Transference Numbers in Mixed Electrolytes.—If a solution contains several ions, produced by a single electrolyte, e.g., sulfuric acid which gives H^+ , HSO_4^- and SO_4^{--} , or by a mixture of electrolytes, each ion will take part in the transport of current. The quantity of electricity (q) carried by any ionic species is proportional to its concentration (c) expressed in g. equiv. per unit volume, and to its velocity (u); hence (cf. p. 895),

$$q = k(cu). \quad (41)$$

The total quantity of electricity (Q) carried by all the ions in the electrolyte is proportional to the sum of the cu terms for all the ions, both positive and negative, so that

$$Q = k(c_1u_1 + c_2u_2 + c_3u_3 + \cdots) = k\Sigma cu. \quad (42)$$

The transference number (t) of any ion is the fraction of the total current carried by that ion, so that by (41) and (42)

$$t_1 = \frac{q_1}{Q} = \frac{c_1u_1}{\Sigma cu}; \quad t_2 = \frac{q_2}{Q} = \frac{c_2u_2}{\Sigma cu}; \quad \text{etc.} \quad (43)$$

Since cu is proportional to the amount of electricity carried by any ion, it is proportional to the number of g. equiv. of that ion migrating from the appro-

priate compartment, whereas Σcu is proportional to the number of g. equiv. deposited in a coulometer. The transference number of an ion in a mixture is, therefore, equal to the number of g. equiv. of that ion migrating from the compartment, divided by the total number of g. equiv. deposited. This principle has been utilized, to some extent, for determining the transference numbers of the different ions in a mixture of two salts with a common ion.

The moving boundary method can also be used for mixtures if one of the ions has a much larger speed than the other, for example as in a mixture of hydrochloric acid and potassium chloride (L. G. Longworth, 1930). Starting with a uniform solution of these two electrolytes and a cadmium anode, two sharp boundaries were observed and studied. After a short time an autogenic boundary formed between cadmium chloride and potassium chloride, and another between the latter and the mixture with hydrochloric acid. From the rates of movement the transference numbers of both cations can be calculated.

Hydration of Ions: "True" Transference Numbers.—It will be seen from Table 151 that the conductances of the alkali metal ions increase with increasing atomic weight, whereas the reverse would have been expected; the smallest ion, lithium, should have the largest velocity, but it actually has the lowest in the series. To explain this discrepancy G. Bredig (1894) suggested that the ions are attached in solution to a number of water molecules, and these "hydrated ions" migrate as a whole under the influence of the applied E.M.F. If the lithium ion is more highly hydrated than sodium, and so on through the series, the size of the ion-hydrate complex may decrease in going from lithium to cesium. In view of the high polarizing power of a small ion (p. 540), it is probable that the attraction between the ion and the molecular dipoles of water will decrease steadily with increasing atomic weight, as required to account for the observed ionic velocities. If the ions do carry molecules of water with them during migration, it is clear that transference number measurements will be affected, since the concentration of the solution may be altered if the two ions are hydrated to different extents.

Suppose each ion has attached to it w_+ and w_- , respectively, molecules of water, and let t_+ and t_- be the *true* transference numbers, if no water were carried by the ions, whereas t'_+ and t'_- are the apparent numbers, as measured. For the passage of one faraday the cations will carry w_+t_+ moles of water in one direction and the anions will carry w_-t_- moles in the opposite direction. There will thus be a net transfer of $w_+t_+ - w_-t_- = x$ moles of water from anode to cathode compartment, and this will account for the difference between the true cation transference number t_+ and the apparent value t'_+ determined by the fall in concentration of the anode compartment. The latter number will be smaller than the former by an amount determined by the quantity of electrolyte associated with the x moles of water, for the concentration of the solution will be increased by this amount. Since it is assumed that one faraday has passed, that is a total of 1 g. equiv. is deposited, the loss of electrolyte from the anode compartment, expressed in g. equiv., is equal to the measured cation transference number, and so the error due to the transport of water is equal to the actual number of g. equiv. of electrolyte associated with the transferred water. If the original solution contained n , g. equiv. of salt together with n_0

moles of water, then $x(n_s/n_w)$ g. equiv. of salt are associated with x moles of water, and this must be equal to the difference between t_+ and t'_+ , so that

$$t_+ - t'_+ = x(n_s/n_w). \quad (44)$$

If the net transfer of water x could be determined experimentally then the true transference number could be evaluated from the apparent value, and knowing this some information could be obtained concerning the hydration numbers w_+ and w_- .

W. Nernst (1900) suggested that the transport of water could be studied by adding to the electrolytic solution an indifferent reference substance, e.g., a sugar, which did not migrate during the passage of current. If no transfer of water occurred the concentration of the reference substance would remain unaltered, but if there were any change the net amount of water could be estimated from this change. The earliest experiments in this connection failed, but E. W. Washburn (1909) was successful, using the sugar raffinose as reference substance, its concentration being derived from its optical rotation. He determined the values of x in 1.25 *N* solutions of lithium, sodium, potassium and cesium chlorides and hydrochloric acid, and hence estimated the true transference numbers. Since x is equal to $w_+t_+ - w_-t_-$, there are still two unknowns, w_+ and w_- , which it is not possible to evaluate without making an assumption. All the salts studied were chlorides, and so the hydration number of the chloride ion was assumed to be either (a) four, or (b) nine, and the others determined from the experimental data for x , w_+ , t_+ and t_- for the given solution. The results for the hydration values, estimated by Washburn for 1.25 *N* solutions, are given in Table 155. In view of the uncertain effect of the raffinose on the

TABLE 155. HYDRATION VALUES FROM TRANSFERENCE DATA

Cl ⁻	H ⁺	Li ⁺	Na ⁺	K ⁺	Cs ⁺
(4)	1.0	14	8.4	5.4	4.7
(9)	2.0	25.3	16.6	10.5	9.9

migration of the ions of the electrolyte the results must be regarded as approximate; they nevertheless show, as expected, that the extent of hydration of the alkali metal ions decreases with increasing atomic weight. Other methods have been proposed for investigating ionic hydration; they need not be discussed here, but it may be stated that the general conclusions are the same as those reached above.

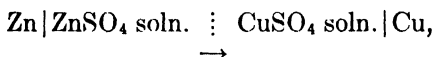
Since there is a transfer of water during electrolysis it is necessary to consider how far transference number measurements are influenced; the maximum difference between true and apparent values observed by Washburn was 0.026 for 1.25 *N* lithium chloride. It is evident from (44) that $t - t'$ is approximately proportional to the concentration of the solution, provided the hydration numbers do not change appreciably; for dilute solutions, therefore, the difference between true and apparent transference numbers may be neglected.¹⁸

THE E.M.F.'S OF CELLS AND ELECTRODE POTENTIALS

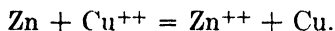
Reversible Cells.—In a galvanic or voltaic cell, or element, chemical energy, or a related form of potential energy, is converted into electrical

energy; such cells generally consist of one or more electrolytic solutions into which dip two electrodes of metal. It will be seen shortly that these electrodes may be of the same or of different materials under different conditions. The electrical energy produced by the cell is measured by the product of its E.M.F. and the quantity of electricity which is passed. If the former is expressed in volts and the latter in coulombs, the electrical energy is in volt-coulombs or joules * (see p. 6), and the equivalent in calories is obtained on multiplying by 0.2390.

In order that the electrical energy may be related thermodynamically to the energy of the process occurring in the cell it is essential that the latter should behave reversibly in the thermodynamic sense (p. 195). A **reversible cell** must satisfy the following conditions: if it is connected to an outside source of E.M.F. exactly equal to that of the cell, then no chemical reaction should occur within it; if the external E.M.F. is decreased by an *infinitesimally* small amount current should flow from the cell, but if the external E.M.F. is increased by an equally small amount the same current should flow through the cell in the opposite direction and any chemical, or other, reaction occurring when the cell produces current should be exactly reversed. A familiar instance of a reversible cell is the Daniell cell, consisting of a zinc electrode immersed in an aqueous solution of zinc sulfate and a copper electrode dipping in copper sulfate solution; the two solutions are generally separated by placing one in a porous pot and the other in the surrounding vessel. The cell may be represented as



the arrow showing the direction in which the positive current flows through the cell when it is producing current. During the passage of current the zinc electrode dissolves and copper is deposited from the solution on to the copper electrode, so that the chemical reaction occurring in the cell is



If an external E.M.F. slightly larger than that of the Daniell cell is applied to it, then current flows in the opposite direction and the cell reaction is reversed. The application of an E.M.F. equal to that of the cell will cause all reaction to cease. It is important to note that the Daniell cell, and similar elements, are reversible only when infinitesimally small currents are passing, for then the system is always in a state of equilibrium as required for strict reversibility (p. 195). Fortunately, in the usual method of determining E.M.F. the current is zero at the instant of measurement, and so this condition for reversibility is satisfied.

Reversible Electrodes.—A reversible cell is always made up of at least two reversible electrodes, or half-elements; such electrodes are of three types. The first consists of a metal dipping into a solution contain-

* The international volt-coulomb, which is identical with the int. joule, is equivalent to 1/4.1833, i.e., 0.2390 (defined) cal. (see p. 184).

ing its ions as, for example, the two electrodes in the Daniell cell; hydrogen may be used to form an electrode of this type (see p. 994). Such electrodes are reversible with respect to the ions of the metal or hydrogen. In the second group the electrodes are reversible with respect to anions; oxygen and halogen electrodes belong to this class, but they are not of great importance. Anion electrodes are best obtained by means of a metal in contact with one of its insoluble salts immersed in a solution of a soluble salt of the same anion, e.g., $\text{Hg}|\text{Hg}_2\text{Cl}_2(\text{s})$ in KCl soln. If current is passed from electrode to electrolyte, mercury dissolves to form mercurous ions; these unite with chloride ions in the solution to form insoluble mercurous chloride, and the result is the removal of chloride ions from solution. On passing current in the opposite direction mercurous ions are discharged, the solution becomes unsaturated and mercurous chloride dissolves; the result is that chloride ions have passed into solution. The system thus behaves as a reversible chlorine electrode, for the chemical changes can be made to occur in either direction by infinitely small changes of applied external E.M.F. Reversible bromine, iodine, oxygen and even sulfate electrodes, e.g., $\text{Hg}|\text{Hg}_2\text{SO}_4(\text{s})$ H_2SO_4 soln., can be obtained in a similar manner. The third type of reversible electrode consists of an unattackable metal, e.g., platinum, in a solution containing ions in two valence states, e.g., Fe^{+++} and Fe^{++} ; the reversible change occurring at the electrode involves conventional oxidation or reduction.

Chemical and Electrical Energy.—In the middle of last century it was generally accepted that the electrical energy of a reversible cell was equal to the chemical energy, measured as heat, of the reaction occurring in the cell; in other words, the electrical energy, determined at constant pressure, was assumed to be equal to the change in heat content of the system. This view was supported by the fact that the electrical energy produced by the Daniell cell when 1 g. atom of zinc dissolves, that is, for the passage of two faradays, is $2 \times 96,500 \times 1.10$ volt-coulombs, where 1.10 volts is the E.M.F. of the cell; multiplication by 0.2390 gives the equivalent of this energy as 50,700 cal. The heat evolved for the reaction between 1 g. atom of zinc and a solution of copper sulfate such as exists in the Daniell cell is about 50,200 cal., in good agreement with the electrical energy. It was realized, however, by J. W. Gibbs (1875), and independently by H. von Helmholtz (1882), that the electrical energy of a reversible cell is equal to the *free energy* change of the process taking place in the cell (cf. p. 921).* If n faradays, i.e., nF coulombs, where F is 96,500, pass through a reversible cell, the change of free energy is equal to nFE , or

$$-\Delta F = nFE, \quad (45)$$

where E is the reversible E.M.F.; the negative sign implies that there is a decrease of free energy, i.e., a spontaneous process is possible, when the

* It is the free energy change ΔF , rather than the change in work content ΔA , because purely mechanical work due to a change in volume is not included (cf. p. 230).

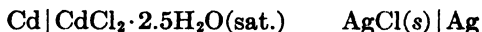
E.M.F. is positive.* The connection between the E.M.F. and the change of heat content (ΔH) can be derived by the appropriate form of the Gibbs-Helmholtz equation (p. 232), viz.,

$$\Delta F = \Delta H + T \left(\frac{\partial(\Delta F)}{\partial T} \right)_P, \quad (46)$$

and utilizing (45) it follows that

$$E = -\frac{\Delta H}{nF} + T \left(\frac{\partial E}{\partial T} \right)_P. \quad (47)$$

Whether the electrical work is greater than, equal to, or less than ΔH depends on the sign of $\partial E/\partial T$, i.e., the temperature coefficient of the E.M.F.† For the Daniell cell this coefficient is very small, and hence it appears, as shown above, that the electrical work is equal to the change of heat content. If $\partial E/\partial T$ is positive, the electrical energy produced will be greater than the heat of the cell reaction at constant pressure; the additional energy is then absorbed from the surroundings, and unless heat is supplied the temperature of the cell will fall during operation. Similarly, if $\partial E/\partial T$ is negative, the difference between nFE and ΔH is given out as heat to the surroundings. It will be seen that if E and $\partial E/\partial T$ are known the heat of reaction of the process occurring in the cell can be evaluated. For example, the E.M.F. of the cell



in which the chemical reaction is



for the passage of two faradays, is 0.6753 volt at 25° and 0.6915 volt at 0° c.; the temperature coefficient $\partial E/\partial T$ over this range is thus -0.00065 volt/degree, at atmospheric pressure. Inserting the value for E at 25° and the temperature coefficient in (47), with n equal to 2, it is found that ΔH is -40,050 cal., compared with -39,530 cal. obtained from thermochemical measurements. The agreement is probably within the limits of experimental error of the latter, and similar or better agreement has been observed in many other instances. The view that the electrical energy, measured in a reversible cell, is equal to the free energy of the reaction occurring may, therefore, be regarded as substantiated. It may be noted, incidentally, that thermal data can frequently be obtained from E.M.F. measurements which are more accurate than those given by direct thermochemical methods.

Measurement of E.M.F.—The principle most commonly employed is that of the Poggendorff compensation method. A cell C (Fig. 237), usually a storage

* For a discussion of the convention for signs of E.M.F. and electrode potentials, see pp. 935, 939.

† In other words, it depends on whether ΔS , the entropy change accompanying the cell reaction, is negative, zero or positive, respectively, since $(\partial(\Delta F)/\partial T)_P$ is equal to $-\Delta S$.

battery, of constant E.M.F. larger than any to be measured, is connected across a uniform wire AB of high resistance. The cell X under examination is connected to A , with the poles in the same direction as the cell C , then through a galvanometer G to a sliding contact D , which can be moved along AB . The

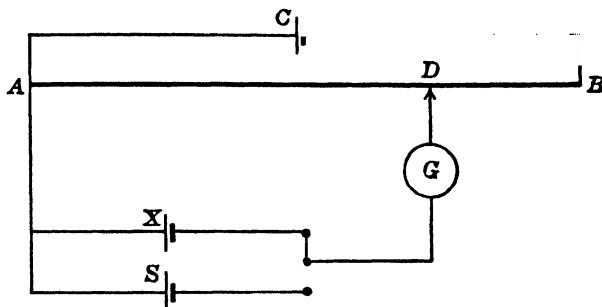


FIG. 237. Measurement of E.M.F.

position of D is so adjusted that no current flows through the galvanometer; the fall of potential along AD due to the cell C is exactly compensated by the E.M.F. of X , that is E_X . By means of a suitable switch the cell X is now replaced by a standard cell, S , of accurately known E.M.F. equal to E_S and the slider re-adjusted until a point of balance is reached at D' ; the fall of potential along AD' is then equal to E_S , and since the wire is uniform

$$E_X = (AD/AD')E_S.$$

It is thus possible to calculate E_X from the measurements, as E_S is known.

The standard cell in common use is the Weston cell; the electrode marked "positive" consists of mercury covered with solid mercurous sulfate and the "negative" electrode is a 12.5 per cent cadmium amalgam; the solution is one of cadmium sulfate kept saturated by crystals of $3\text{CdSO}_4 \cdot 8\text{H}_2\text{O}$ placed over the cadmium electrode. The E.M.F. of this cell in international volts may be calculated for any temperature $t^\circ\text{C}$. by the formula

$$E = 1.01830 - 4.06 \times 10^{-5}(t - 20) - 9.5 \times 10^{-7}(t - 20)^2 + 1 \times 10^{-9}(t - 20)^3.$$

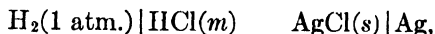
The temperature coefficient is small enough to be neglected and at normal temperatures the value 1.0183 volts, for 20°C ., may be adopted. As a secondary standard, for general laboratory purposes, the "unsaturated" Weston cell is often employed; it contains a solution of cadmium sulfate saturated at 4°C ., so that it is unsaturated at room temperatures. The E.M.F. of this cell is 1.0186 volts, and it is virtually independent of the temperature under normal conditions.

For approximate measurements the potentiometer wire AB may be a straight or spiral length of platinum, platinum-iridium or other resistant metal. The position of the sliding contact can usually be read with an accuracy of 0.5 mm., and this corresponds to an error of 0.001 volt if C is 2 volts and AB is 1 meter long. For accurate work the wire may be replaced by two calibrated resistance boxes; the contact D is fixed where the two boxes are joined and the potential across AD is varied by changing the resistances in the boxes, keeping

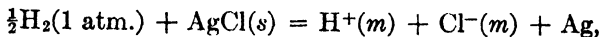
the total constant. A number of special potentiometers are available commercially which operate on the Poggendorff principle; the E.M.F. can be read directly with an accuracy of 0.1 millivolt. If approximate data only are required, *AB* may be replaced by a nonuniform wire or by suitable rheostats. When the system is adjusted so that there is no current through the galvanometer, the fall of potential across *AD* is measured by a voltmeter. This procedure is known as the "potentiometer-voltmeter" method; it has been employed in various forms for electroanalytical work.

As the balance point detector a suitably damped mirror galvanometer, with a high megohm sensitivity, should be used for accurate work, but for approximate purposes a simple pointer galvanometer is often satisfactory. The capillary electrometer, the action of which depends on the change of tension at a mercury-sulfuric acid interface with applied potential (cf. p. 1229), at one time widely employed in E.M.F. work, is now rarely used; it has the advantages of being unaffected by electrical and magnetic disturbances, and by high resistances, but it is relatively insensitive and is liable to behave erratically in damp weather. If the cell to be measured has a very high resistance, a quadrant electrometer, or a ballistic galvanometer in conjunction with a condenser which is charged by the cell and discharged through the galvanometer, may be used to indicate the balance point. It is generally more convenient, however, to employ a form of vacuum-tube potentiometer; the principle involved is roughly as follows. A potentiometer from which a known E.M.F. can be tapped off, e.g., by the use of a standard cell, is placed between the filament and grid of the tube, and the E.M.F. is adjusted so as to give a suitable anode current; this is then either noted or exactly compensated to zero by a subsidiary current. The unknown cell is now included with the potentiometer in the grid-filament circuit, and the potentiometer is adjusted until the anode current is the same as before. The difference in the two readings of the potentiometer gives the E.M.F. of the cell. On account of the high impedance of the tube, cells of resistance up to 10^6 ohms can be measured without difficulty. For higher resistances special tubes or special forms of circuit have been devised.¹⁷

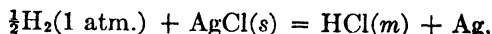
Concentration Cells without Transference.—A concentration cell is a galvanic cell in which the electrical energy is derived from the free energy change accompanying the transfer of a substance from a system of high to one of low concentration; these cells are the simplest from the thermodynamic standpoint, for they involve no *net* chemical reaction. Consider the cell



in which the chemical reaction (cf. p. 934)



or more conventionally,



takes place for the passage of one faraday; *m* is the molality (p. 636) of the solution of hydrochloric acid, upon which the E.M.F. of the cell depends. The special characteristic of this cell is that each electrode is reversible with respect to one of the ions of the electrolyte; the left-hand

one with respect to hydrogen and the right with respect to chloride ions. If two of these cells, containing hydrochloric acid solutions m_1 and m_2 and having E.M.F.'s E_1 and E_2 , respectively, are connected so as to oppose each other, the cell



with an E.M.F. of $E_2 - E_1$ will be obtained. Since both cells are reversible and they operate in different directions, it is at once evident that there can be no *resultant* chemical reaction. If m_1 is greater than m_2 , then in the left-hand cell hydrogen gas reacts with solid silver chloride to produce silver and hydrochloric acid, in the form of hydrogen and chloride ions, whereas in the right-hand cell the acid and silver unite to re-form hydrogen and solid silver chloride. The net result is the transfer of hydrogen gas at 1 atm. from the left-hand to the right-hand electrode, of solid silver chloride from left to right, and of hydrochloric acid from molality m_1 to m_2 . Of these three transfers the first two clearly involve no change of free energy, since the chemical potentials of both hydrogen and silver chloride are unaltered. There is, however, a change in the free energy of the hydrogen and chloride ions constituting the hydrochloric acid. The free energy change for the transfer of 1 g. ion of hydrogen ions, at constant temperature and pressure, is

$$\Delta F_{\text{H}^+} = (\mu_{\text{H}^+})_2 - (\mu_{\text{H}^+})_1 = RT \ln (a_{\text{H}^+})_2 / (a_{\text{H}^+})_1, \quad (48)$$

where μ_{H^+} and a_{H^+} are the chemical potentials and activities, respectively, of the hydrogen ions in the solutions indicated by the subscripts 1 and 2. Similarly, the change of free energy accompanying the transfer of 1 g. ion of chloride ions is

$$\Delta F_{\text{Cl}^-} = (\mu_{\text{Cl}^-})_2 - (\mu_{\text{Cl}^-})_1 = RT \ln (a_{\text{Cl}^-})_2 / (a_{\text{Cl}^-})_1, \quad (49)$$

and addition of (48) and (49) gives the total free energy change, viz.,

$$\Delta F = RT \ln \frac{(a_{\text{H}^+})_2 (a_{\text{Cl}^-})_2}{(a_{\text{H}^+})_1 (a_{\text{Cl}^-})_1} = 2RT \ln \frac{a_2}{a_1}, \quad (50)$$

where a_1 and a_2 are the *mean* activities of the hydrochloric acid in the two solutions; the mean activity of an electrolyte is defined as the *geometric mean of the activities of its constituent ions* (p. 959). The electrical energy produced in the cell for the passage of one faraday is FE , where E is $E_2 - E_1$; as seen above this is equal to $-\Delta F$, so that

$$E = E_2 - E_1 = \frac{2RT}{F} \ln \frac{a_1}{a_2}. \quad (51)$$

Replacing the mean activity (a) by the product of the molality (m) and the mean activity coefficient (γ),* it follows that

$$E = \frac{2RT}{F} \ln \frac{m_1 \gamma_1}{m_2 \gamma_2}. \quad (52)$$

* It is the common practice to use the symbol γ for the activity coefficient f_m , as defined on p. 685.

In dilute solutions the ratio of the activity coefficients is approximately unity, and so

$$E \approx \frac{2RT}{F} \ln \frac{m_1}{m_2} \approx \frac{2RT}{F} \ln \frac{c_1}{c_2}, \quad (53)$$

and hence the E.M.F. of the complete cell depends on the ratio of the concentrations of the electrolyte in the two parts. A cell of the type under consideration is called a **concentration cell without transference**, for there is no *direct* transfer of electrolyte from one solution to the other. A cell of this description is obtained whenever two simple cells of the type described on p. 926, i.e., with each electrode reversible with respect to one of the ions of the electrolyte, containing solutions of different concentrations are combined. The direction in which the cells are connected is immaterial, as long as they are in opposition; for example, the two cells already considered may be combined to give



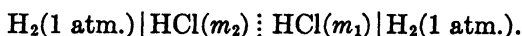
As is to be expected, the E.M.F. is found to be identical numerically with that for the alternative combination, as given by (51); the flow of current is, however, in the opposite direction and this should be indicated by a negative sign or by inverting the activity fraction.

The general equation for the E.M.F. of any concentration cell without transference can be derived in the manner described above; it is found to be

$$E = \pm \frac{\nu}{\nu_{\pm}} \cdot \frac{RT}{z_{\pm}F} \ln \frac{a_1}{a_2}, \quad (54)$$

where ν is the total number and ν_+ or ν_- the number of positive or negative ions produced by the ionization of one molecule of electrolyte, and z_+ or z_- is the valence of the ion with respect to which the extreme electrodes are reversible. If this ion is positive, then the positive signs apply throughout; if negative, then the negative signs are applicable.*

Concentration Cells with Transference.—If the intermediate electrode, e.g., $\text{AgCl}(s)|\text{Ag}|\text{AgCl}(s)$, is removed from the concentration cell without transference which has been considered, there remains the cell



In this cell the two solutions are in contact and there is direct transfer of electrolyte from the more concentrated to the more dilute solution. In general, whenever two solutions of an electrolyte at different concentrations are brought together, and the same electrode, reversible with respect to one of the ions of the electrolyte, is placed in each solution, there is obtained a **concentration cell with transference**. When one faraday passes through the cell shown, 1 g. atom of hydrogen dissolves at the left-hand electrode, where m_2 is the more dilute of the two solutions, to

* It may be noted that ν_+z_+ is equal to ν_-z_- .

give 1 g. ion of hydrogen ions. At the same time t_{H^+} g. ion of hydrogen ions will migrate across the boundary in the direction of the current, i.e., from left to right, and t_{Cl^-} g. ion of chloride ions will pass in the opposite direction; the quantities t_{H^+} and t_{Cl^-} are the transference numbers of the two ions concerned. Finally, 1 g. ion of hydrogen ions will be discharged at the right-hand electrode to produce 1 g. atom of hydrogen. The net result of the passage of one faraday is the transfer of $1 - t_{H^+} = t_{Cl^-}$ g. ion of hydrogen ions, and t_{Cl^-} g. ion of chloride ions from right to left, i.e., from m_1 to m_2 . By the same method as that adopted for the concentration cell without transference, the total change of free energy is found to be

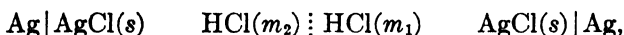
$$\Delta F = t_{Cl^-} RT \ln \frac{(a_{H^+})_2 (a_{Cl^-})_2}{(a_{H^+})_1 (a_{Cl^-})_1} = 2t_{Cl^-} RT \ln \frac{a_2}{a_1}, \quad (55)$$

and since the electrical work is FE , which is equal to $-\Delta F$,

$$E = 2t_{Cl^-} \frac{RT}{F} \ln \frac{a_1}{a_2} = 2t_{Cl^-} \frac{RT}{F} \ln \frac{m_1 \gamma_1}{m_2 \gamma_2}. \quad (56)$$

The E.M.F. of the concentration cell with transference is thus seen to involve the transference number, in addition to the activities, or concentrations, of the two solutions.

The electrodes of a concentration cell may be reversible with respect to either of the ions of the electrolyte. In the cell just considered the electrodes are reversible with respect to hydrogen ions, but in the form



the electrodes are reversible with respect to chloride ions. The current now flows in the opposite direction, and for the passage of one faraday 1 g. atom of silver dissolves at the right-hand electrode to form 1 g. ion in solution, which unites with 1 g. ion of chloride ions to form solid silver chloride. At the same time the reverse process occurs at the left-hand electrode, while t_{H^+} g. ion of hydrogen ions migrate in the direction of the current and t_{Cl^-} g. ion of chloride ions pass in the opposite direction. The net result is the transfer of t_{H^+} g. ion of hydrogen and of chloride ions from m_1 to m_2 , and the E.M.F. of the cell is given by

$$E = 2t_{H^+} \frac{RT}{F} \ln \frac{a_2}{a_1} = 2t_{H^+} \frac{RT}{F} \ln \frac{m_2 \gamma_2}{m_1 \gamma_1}. \quad (57)$$

It will be noted that the activity fraction has been reversed to show that the flow of current is in the opposite direction to that of the cell with hydrogen electrodes. Further, the E.M.F. is now dependent on the transference number of the hydrogen ions. In general, the E.M.F. of any cell with transference can be written in the form

$$E = \pm t_{\mp} \frac{\nu}{z_{\pm}} \frac{RT}{F} \ln \frac{a_1}{a_2}, \quad (58)$$

where the symbols have the same significance as in (54).

It should be noted that, in the above deductions it has been assumed that the transference numbers are independent of concentration; this may be justifiable if the two solutions do not have very different concentrations. For two solutions of activity $a + da$ and a , respectively, the E.M.F., represented by dE , is given by

$$dE = 2t_{Cl^-} \frac{RT}{F} d \ln a, \quad (59)$$

for the hydrogen-hydrochloric acid cell, utilizing the fact that $\ln(a + da)/a \approx d \ln a$. The E.M.F. for the cell with solutions m_1 and m_2 is obtained by integration; thus,

$$E = 2 \frac{RT}{F} \int_{a_2}^{a_1} t_{Cl^-} d \ln a. \quad (60)$$

If t_{Cl^-} may be assumed constant then (60) becomes identical with (56).

Transference (Transport) Numbers from Concentration Cells.—If E_t is the E.M.F. of a concentration cell with transference, and E is that of the corresponding cell without transference, then it is evident from (54) and (58) that

$$E_t/E = t_{\mp}, \quad (61)$$

so that the ratio of the E.M.F.'s gives the transference number of the ion other than that with respect to which the extreme electrodes are reversible. It should thus be possible to evaluate the transference number of an ion from two E.M.F. measurements; this would only be feasible, however, if the transference number were independent of concentration.

Consider two cells in which the electrolyte concentrations differ very little, so that the corresponding activities are a and $a + da$; then [cf. equation (59)]

$$dE_t = \pm t_{\mp} \frac{\nu}{\nu_{\pm}} \cdot \frac{RT}{z_{\pm} F} d \ln a \quad (62)$$

and

$$dE = \pm \frac{\nu}{\nu_{\pm}} \cdot \frac{RT}{z_{\pm} F} d \ln a; \quad (63)$$

$$\therefore \frac{dE_t}{d \ln a} \bigg/ \frac{dE}{d \ln a} = t_{\mp}. \quad (64)$$

If the E.M.F.'s of cells, with and without transference, in which the concentration of one of the solutions is varied and the other kept at a constant low value, e.g., 0.001M, are plotted against $\ln a$, or $\log a$, of the variable solution, the slopes of the curves are $dE_t/d \ln a$ and $dE/d \ln a$, respectively. The transference number may thus be determined at any concentration, from the slopes at the value of a corresponding to this concentration. The activity at any particular concentration must be determined independently by E.M.F. or other methods. In view of the difficulty in the precise measurement of the slopes of the curves mathematical devices

have been used by D. A. MacInnes and J. A. Beattie (1920), by G. Jones and M. Dole (1929), and others. The results are in satisfactory agreement with those of the methods described previously.¹⁸

Liquid Junction Potentials.—When two solutions of the same electrolyte are in contact, as in cells with transference, the more concentrated solution will tend to diffuse into the more dilute; the rate of diffusion of each ion is roughly proportional to its velocity in an electric field. If the cation moves more rapidly than the anion, the former will tend to diffuse ahead of the latter into the dilute solution; this solution will thus become positively charged with respect to the concentrated solution. Should the anion be the faster moving ion, the dilute solution will acquire a negative charge. In either case an electrical "double layer" is produced at the junction of the solutions, and the attraction between opposite charges prevents any appreciable separation of anions from cations; there must, however, be a difference of potential at the boundary, and this is called the **liquid junction potential**, or sometimes the **diffusion potential**. Its magnitude depends on the relative speeds of the ions, and its direction is such that the more dilute solution always assumes the charge of the faster moving ion.

In the hydrogen-hydrochloric acid cell considered above, two solutions of hydrochloric acid at different concentrations are in contact, and so there will be a liquid junction potential which may be represented by E_l . If one faraday of electricity passes through the cell then, as already seen, t_{H^+} g. ion of hydrogen ions migrate from m_2 to m_1 , while t_{Cl^-} g. ion of chloride ions move from m_1 to m_2 ; the total change of free energy is

$$\Delta F = t_{H^+} RT \ln \frac{(a_{H^+})_1}{(a_{H^+})_2} + t_{Cl^-} RT \ln \frac{(a_{Cl^-})_2}{(a_{Cl^-})_1}. \quad (65)$$

The electrical work $E_l F$ is equal to $-\Delta F$, and hence

$$E_l = t_{H^+} \frac{RT}{F} \ln \frac{(a_{H^+})_2}{(a_{H^+})_1} + t_{Cl^-} \frac{RT}{F} \ln \frac{(a_{Cl^-})_1}{(a_{Cl^-})_2}. \quad (66)$$

Since t_{H^+} is equal to $1 - t_{Cl^-}$, and remembering that the mean activity is the square root of $a_{H^+} \times a_{Cl^-}$, it follows that

$$E_l = 2t_{Cl^-} \frac{RT}{F} \ln \frac{a_1}{a_2} - \frac{RT}{F} \ln \frac{(a_{H^+})_1}{(a_{H^+})_2}. \quad (67)$$

This is the correct form of the equation for the liquid junction potential, assuming constant transference numbers, but for approximate purposes it may be simplified. If the ratio of the activities is the same for both ions, then $(a_{H^+})_1/(a_{H^+})_2$ is equal to a_1/a_2 , and (67) becomes

$$E_l \approx (2t_{Cl^-} - 1) \frac{RT}{F} \ln \frac{a_1}{a_2}. \quad (68)$$

The general equation, of *this form*, for the liquid junction potential is

$$E_l \approx \left(t_- \frac{\nu}{\nu_+} - 1 \right) \frac{RT}{z_+ F} \ln \frac{a_1}{a_2}, \quad (69)$$

where the symbols and signs have the same meaning as before. For a simple uni-univalent electrolyte (69) reduces to the form of (68). It is seen that the sign of the liquid junction potential depends on the relative speeds of the ions; if the cation is the faster then, for a binary electrolyte, t_- will be less than 0.5, and E_l will have a negative sign, but if the anion is the faster, t_- is greater than 0.5 and the sign is positive.

Electrode Concentration Cells.—Concentration cells may be made up of two electrodes of the same metal, or other material, at different concentrations immersed in a solution containing ions of the metal. The amalgam cells, consisting of two different amalgams of the same metal, e.g., zinc, copper, lead, tin or cadmium, and a solution of one of the salts of the metal, are examples of this type of concentration cell. Suppose the activities of the metal in the two amalgams, of concentrations c_1 and c_2 , are a_1 and a_2 ; then the removal of 1 mole of metal from the more concentrated amalgam (c_1) and its deposition in the dilute amalgam (c_2) requires mzF coulombs, where m is the number of atoms in the molecule and z is the valence of the ion. The electrical work, equal to $-\Delta F$, is thus $mzFE$, where E is the E.M.F. of the cell. The free energy change accompanying the transfer of 1 mole of metal from activity a_1 to a_2 is

$$\Delta F = \mu_2 - \mu_1 = RT \ln a_2/a_1$$

at constant temperature and pressure; hence

$$E = \frac{RT}{mzF} \ln \frac{a_1}{a_2}. \quad (70)$$

For zinc and copper m is 1 and z is 2, so that

$$E = \frac{RT}{2F} \ln \frac{a_1}{a_2} \approx \frac{RT}{2F} \ln \frac{c_1}{c_2}, \quad (71)$$

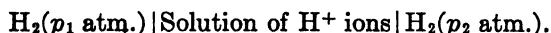
the ratio a_1/a_2 being replaced by c_1/c_2 in the approximate form; the latter has been found to represent the behavior of cells containing dilute amalgams.

Equations (70) and (71) do not include the sign of the E.M.F., but this may be readily deduced: the more concentrated amalgam will clearly have the greater tendency to send positive ions into solution, and so, in accordance with convention (p. 935), it will give the more positive electrode.

Other cells of a similar type may be made up with electrodes consisting of the pure metal and an amalgam, or with the same metal in two different conditions, e.g., allotropic forms, crystals of different sizes, or normal and

strained states. Gravity cells, in which the electrodes, e.g., of mercury, consist of columns of unequal length belong to the same category.

An interesting concentration cell consists of two hydrogen gas electrodes at different pressures; thus,



If p_1 is the higher pressure than this electrode will have the more positive potential, and the passage of two faradays will be accompanied by the dissolution of 1 mole of hydrogen from the left-hand electrode, i.e., at the higher pressure, to form 2 g. ions of hydrogen ions in solution; at the same time 2 g. ions will be discharged to produce 1 mole of hydrogen at the right-hand electrode. Assuming hydrogen to behave as an ideal gas, it follows that at constant temperature

$$2FE_{\text{H}_2} = -\Delta F = RT \ln p_1/p_2; \quad (72)$$

$$\therefore E_{\text{H}_2} = \frac{RT}{2F} \ln \frac{p_1}{p_2}. \quad (73)$$

To include the sign of the E.M.F. this should be preceded by a positive sign, to show that the left-hand electrode is the positive one. The reliability of (73) has been confirmed over a range of pressures from 0.005 to 400 atm.; at higher pressures deviations occur because of departure of the gas from ideal behavior.

Theoretically an oxygen cell of the same kind is possible; since the dissolution and deposition of 1 mole of oxygen requires four faradays, the oxygen atom being bivalent, the electrical work is $4FE$. The free energy decrease is $RT \ln p_1/p_2$ per mole, as before, and so

$$-E_{\text{O}_2} = \frac{RT}{4F} \ln \frac{p_1}{p_2}, \quad (74)$$

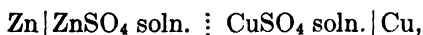
the negative sign being used to show that the direction of the cell is opposite to that with hydrogen. The oxygen electrode does not behave in a reversible manner, and so the verification of (93) is not possible; it has nevertheless important applications in physical chemistry.¹⁹

Concentration Cells and Equilibrium.—All concentration cells have a property in common: when they produce current, matter is always transferred from the more concentrated to the more dilute electrode or solution. In every case, therefore, the two concentrations tend to equalize as the cell operates, and so the E.M.F. falls until it becomes zero when the concentrations are equal. Under these conditions the system is, of course, at equilibrium and since the free energy must then be a minimum the cell must have an E.M.F. of zero.

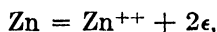
The fact that at equilibrium a concentration cell has no E.M.F. has been utilized to determine transition points. If a cell is made up of two allotropic forms of a metal, e.g., tin, in a solution of one of its salts, there will be an E.M.F., provided the temperature is not the transition point.

Only at this temperature are the two forms in equilibrium and the E.M.F. will then be zero; by varying the temperature of the cell, therefore, the transition point can be determined. Two salt hydrates have the same solubility at the transition point, but not at other temperatures; a cell consisting of saturated solutions of the two hydrates, with electrodes of the metal whose ions are present in the salt, will consequently have no E.M.F. at the transition temperature.

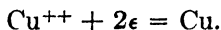
Chemical Reversible Cells.—In the cells considered hitherto the two electrodes are essentially the same, although the concentrations differ; the chemical change at one electrode is the reverse of that at the other, so that there is no over-all chemical reaction. When two different reversible electrodes are combined to form a cell, however, there is a resultant chemical change; the electrical energy obtained is equivalent to the decrease in free energy accompanying the reaction taking place in the cell. Before proceeding to develop the equation for the E.M.F. of a chemical cell, it is necessary to consider the nature of the cell reaction. In the familiar Daniell cell, viz.,



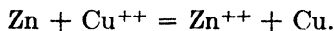
for example, the metallic zinc at the left-hand electrode passes into solution as zinc ions, liberating two electrons, i.e.,



while at the right-hand electrode cupric ions are discharged to form metallic copper, i.e.,

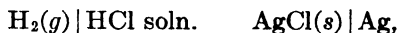


The electrons set free at the zinc electrode may be supposed to travel along the external connecting circuit, and are available for the discharge of the copper ions. The complete cell reaction, obtained by adding the separate electrode reactions, is consequently

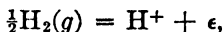


Since two electrons are involved for each atom taking part in the process, the whole reaction as written, with quantities in g. atoms and g. ions, would take place for the passage of two faradays of electricity through the cell.

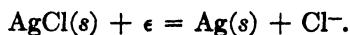
In the reversible cell (cf. p. 926)



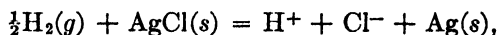
the hydrogen gas forms hydrogen ions in solution, i.e.,



at the left-hand electrode, while at the right-hand, the reaction is

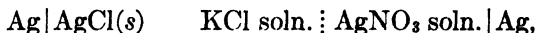


The over-all chemical change is then

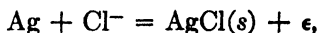


the H^+ and Cl^- implying hydrochloric acid in aqueous solution; the reaction as written would take place for the passage of one faraday.

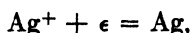
A different type of cell is



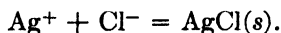
in which both electrodes are of the same metal. At the left-hand electrode, which is the reverse of the right-hand electrode of the preceding cell, the reaction is



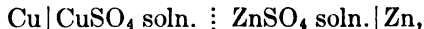
while at the right-hand electrode,



so that the complete cell reaction, for one faraday, is



It will be observed that in each case the process at the left-hand electrode is always accompanied by the liberation of electrons, while the same number are removed in the reaction at the right-hand electrode. In other words, the left-hand electrode process is one of oxidation, using the term in its most general sense, whereas a reduction process takes place at the right-hand electrode. It is true of all reversible chemical cells that there is oxidation at one electrode and reduction at the other. According to the widely adopted convention, the E.M.F. of a cell is taken as *positive when the spontaneous cell reaction is such as to produce oxidation at the left-hand and reduction at the right-hand electrode*. The same convention is sometimes stated in another form: a positive E.M.F. means the spontaneous passage of positive ions through the cell from left to right, and of negative ions from right to left, when the cell operates. If the reverse is true, then according to the convention the E.M.F. would be negative. The cells considered above were deliberately written in such a manner as to give a positive E.M.F., but if they had been set up in a reverse manner, e.g.,



the E.M.F.'s would have been the same numerically, but opposite in sign. By adopting the foregoing conventions, the value of nFE , where n is the number of faradays required for the cell reaction, is equal to the *decrease* of free energy accompanying the process [cf. equation (45)].*

* The "practical" convention, employed in connection with cells for producing current, e.g., the Daniell cell, Leclanché cell, secondary cells, etc., is to call the "negative" the electrode at which the process is oxidation when the cell is giving current; the "positive" is the electrode at which reduction is the spontaneous process. The reason for this is that oxidation is accompanied by the liberation of electrons, and so the elec-

E.M.F. of a Chemical Cell.—For the general reaction



the free energy change is given by the reaction isotherm (p. 827) as

$$\Delta F = \Delta F^0 + RT \ln \frac{(a_L)^{\nu_L} (a_M)^{\nu_M} \dots}{(a_A)^{\nu_A} (a_B)^{\nu_B} \dots}, \quad (75)$$

where the activities in the second term on the right-hand side refer to the *arbitrary* values. If the reaction, with these particular activities, takes place in a reversible chemical cell of E.M.F. equal to E , for the passage of n faradays, then $-\Delta F$ is also equal to nFE . If all the substances contained in the cell are in their respective standard states, that is to say, they are all at unity activity, the E.M.F. is the standard value E^0 for the given cell; in these circumstances the standard free energy decrease $-\Delta F^0$ would be equal to nFE^0 . Combining these results with (75), it is seen that

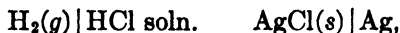
$$E = E^0 - \frac{RT}{nF} \ln \frac{(a_L)^{\nu_L} (a_M)^{\nu_M} \dots}{(a_A)^{\nu_A} (a_B)^{\nu_B} \dots}, \quad (76)$$

which is the general expression for the E.M.F. of any reversible cell. The standard E.M.F., i.e., E^0 , is a constant for the given cell, at a definite temperature. It has an important significance, since it is equal to $-\Delta F^0/nF$, while ΔF^0 is equivalent to $-RT \ln K$, where K is the equilibrium constant of the cell reaction; hence,

$$E^0 = \frac{RT}{nF} \ln K. \quad (77)$$

The standard E.M.F. is thus related to the equilibrium constant of the reaction taking place in the cell. The actual E.M.F., as seen by (76), is also dependent on the activities, i.e., concentrations, etc., of the various substances involved in the cell reaction; if these should all happen to be unity, the E.M.F. would, of course, be equal to the standard value.

The use of (76) may be illustrated by reference to the cell



considered above. From the cell reaction already given it is seen that

$$E = E^0 - \frac{RT}{F} \ln \frac{a_{\text{H}^+} a_{\text{Cl}^-} a_{\text{Ag}}}{a_{\text{H}_2} a_{\text{AgCl}}}, \quad (78)$$

where the various activities refer to the actual values in the cell as set up; E^0 is the standard E.M.F. for the H_2 , $\text{HCl}-\text{AgCl}$, Ag cells. If the

trode metal acquires a "negative" charge; similarly, the reduction electrode will acquire a "positive" charge, because electrons are taken up from it. According to the convention employed in this book, the E.M.F. of a cell is positive when it is set up in such a way that the "negative," i.e., oxidation, electrode is to the left, and the "positive," i.e., reduction, electrode is to the right.

hydrogen gas is at 1 atm. pressure it will be in its conventional standard state, and since the silver and silver chlorides are solids, they are also in their respective standard states; hence a_{H_2} , a_{Ag} and a_{AgCl} are unity, and (78) simplifies to

$$E = E^0 - \frac{RT}{F} \ln a_{H^+} a_{Cl^-} = E^0 - \frac{2RT}{F} \ln a_{HCl}, \quad (79)$$

where a_{HCl} represents the mean activity of the ions in the hydrochloric acid solution. This result is in agreement with the expression derived on p. 927 for the E.M.F. of a concentration cell involving two cells of the type under consideration. By the use of the method described it is thus a simple matter to derive an equation relating the E.M.F. of a reversible chemical cell to the activities of the constituent substances, provided the cell reaction is known.

Single Electrode Potentials.—Since the complete reaction taking place in the cell is made up of the two individual electrode reactions, it is possible to split up the general equation (76) into two parts, each of which is characteristic of a single electrode. This may be illustrated by writing (78) as

$$E = \left(E_H^0 - \frac{RT}{F} \ln \frac{a_{H^+}}{a_{H_2}^{\frac{1}{2}}} \right) - \left(E_{AgCl}^0 - \frac{RT}{F} \ln \frac{a_{AgCl}}{a_{Ag} a_{Cl^-}} \right), \quad (80)$$

where the first term in parentheses on the right-hand side refers to the hydrogen (left-hand) electrode and the second term applies to the silver-silver chloride (right-hand) electrode of the cell. These expressions may thus be identified with the potentials of the separate electrodes, so that

$$E_H = E_H^0 - \frac{RT}{F} \ln \frac{a_{H^+}}{a_{H_2}^{\frac{1}{2}}} \quad \text{and} \quad E_{AgCl} = E_{AgCl}^0 - \frac{RT}{F} \ln \frac{a_{AgCl}}{a_{Ag} a_{Cl^-}}, \quad (81)$$

where E_H^0 and E_{AgCl}^0 are the respective **standard electrode potentials**. If, as above, the hydrogen is at 1 atm. pressure, the activities of the hydrogen, silver and silver chloride may be set equal to unity, so that the potentials in (81) may be simplified to

$$(i) \ E_H = E_H^0 - \frac{RT}{F} \ln a_{H^+} \quad \text{and} \quad (ii) \ E_{AgCl} = E_{AgCl}^0 + \frac{RT}{F} \ln a_{Cl^-}. \quad (82)$$

Combination of these two equations gives (79); it is seen that the standard E.M.F. of the complete cell is equal to the (algebraic) difference of the standard potentials of the individual electrodes.

The foregoing treatment may be generalized by considering the *electrode reaction*



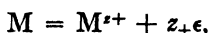
in which n electrons (or n faradays) are involved; the expression for the

electrode potential is then

$$E = E^0 - \frac{RT}{nF} \ln \frac{(a_X)^{r_X}(a_Y)^{r_Y} \dots}{(a_A)^{r_A}(a_B)^{r_B} \dots} \quad (83)$$

It can be readily shown that (83) leads to the equations (81) when applied to the particular electrode reactions concerned. Further, by omitting all factors referring to gases at 1 atm. and to solids, since the corresponding activities are unity, expressions analogous to (82) are obtained.

If the reversible electrode consists of a pure metal M of valence z_+ in contact with a solution of M^{z+} ions, the electrode reaction is



and since the pure metal M is in its standard state, the electrode potential is given by (83) as

$$E_+ = E^0 - \frac{RT}{z_+F} \ln a_{M^{z+}}, \quad (84)$$

where $a_{M^{z+}}$ is the activity of the M^{z+} ions in the solution. The same equation is applicable to a hydrogen electrode, with gas at 1 atm. pressure, since this is in equilibrium with hydrogen (positive) ions in solution [cf. equation (82, i)].

For an electrode involving a substance A (or substances, e.g., Ag , $AgCl$) which is reversible with respect to the anions A^{z-} , the electrode reaction is



and if A is in its standard state, e.g., gas at 1 atm. pressure or solid, the equation for the electrode potential is

$$E_- = E^0 + \frac{RT}{z_-F} \ln a_{A^{z-}}. \quad (85)$$

This result is applicable to oxygen and halogen electrodes [cf. equation (82, ii)], since they are reversible with respect to negative ions.

The general form of (84) and (85), for any electrode reversible with respect to a single ion of valence z_{\pm} , is readily seen to be

$$E_{\pm} = E^0 \mp \frac{RT}{z_{\pm}F} \ln a_i, \quad (86)$$

where a_i is the activity of the particular ionic species in the electrode solution. In (86) the upper signs apply throughout for a positive ion, while the lower signs are used for a negative ion.

Another type of reversible electrode, mentioned briefly on p. 923, involves ions originating from two different valence states of the same element, e.g., Fe^{++} and Fe^{+++} ; the electrode reaction for the ferrous-ferric system is



and hence by (83), the potential is given by

$$E_{\text{Fe}^{++}, \text{Fe}^{+++}} = E^0 - \frac{RT}{F} \ln \frac{a_{\text{Fe}^{+++}}}{a_{\text{Fe}^{++}}}. \quad (87)$$

Electrodes of this type are sometimes said to involve an **oxidation-reduction** system, although as already seen all electrodes involve oxidation or reduction processes in the most general sense. Oxidation-reduction systems can consist of anions, e.g., $\text{Fe}(\text{CN})_6^{4-}$, $\text{Fe}(\text{CN})_6^{3-}$, as well as cations, or both may take part in the electrode reaction. A case of particular interest is the manganous-permanganate system, for which the reaction is



By (83) the electrode potential would be represented by

$$E_{\text{Mn}^{++}, \text{MnO}_4^-} = E^0 - \frac{RT}{5F} \ln \frac{a_{\text{MnO}_4^-} a_{\text{H}^+}^8}{a_{\text{Mn}^{++}}}, \quad (88)$$

the activity of the water being taken as unity, although this is justifiable only for pure water or a very dilute aqueous solution. Other types of oxidation-reduction electrodes, containing organic compounds, will be described later (p. 951).

Sign of Electrode Potential: Oxidation and Reduction Potentials.—

It will be noted that when writing the electrode reactions in the preceding section the equation has been so arranged that the electrons appear on the right-hand side. In other words, the reduced state of the electrode system is on the left and the oxidized state on the right of the equation; the terms oxidized and reduced are here used in their most general significance, e.g., M^+ and A would be the oxidized forms corresponding to the reduced forms M and A^- , respectively. It is thus possible to write the electrode reaction as

Reduced State = Oxidized State + n Electrons,

and hence (83) may be expressed as

$$E = E_0 - \frac{RT}{nF} \ln \frac{(\text{Oxidized State})}{(\text{Reduced State})}, \quad (89)$$

where the parentheses imply the products of the various activity terms. The reaction as written represents an oxidation process, and nFE , which is equal to the decrease of free energy, is a measure of the tendency of this oxidation to take place. For this reason, electrode potentials given by (83) or (89) are referred to as **oxidation potentials**. It has been seen that in every reversible cell an oxidation process occurs at one electrode, the left-hand by convention, and a reduction takes place at the other. The foregoing equations thus give the potential of the left-hand electrode of the cell as usually written. The potential at the other electrode is that

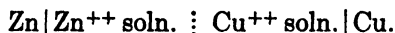
for the reduction process, viz.,

$$\text{Oxidized State} + n \text{ Electrons} = \text{Reduced State},$$

and the corresponding value, known as the **reduction potential**, is the same numerically as the oxidation potential for the same system, but is *opposite in sign*.

In order to facilitate the representation of electrodes a simple convention is adopted: when the electrode is a metal M and the process is oxidation to M^+ ions, the reduced state of the system is written to the left and the oxidized state to the right, viz., M, M^+ , as in the electrochemical equation $M \rightarrow M^+ + \text{electrons}$. Examples of oxidation electrodes are thus $\text{Cu}, \text{Cu}^{++}$ (or $\text{Cu}, \text{CuSO}_4 \text{ soln.}$); $\text{Zn}, \text{Zn}^{++}$ (or $\text{Zn}, \text{ZnSO}_4 \text{ soln.}$); and Ag, Ag^+ (or $\text{Ag}, \text{AgNO}_3 \text{ soln.}$). The potentials of such electrodes are given by (83) or (89). On the other hand, if the electrodes are represented in the reverse manner, i.e., M^+, M , with the oxidized state to the left and the reduced state to the right, as in $\text{Zn}^{++}, \text{Zn}$; $\text{Cu}^{++}, \text{Cu}$; or Ag^+, Ag , the electrode process is reduction, and the potentials are opposite in sign to those for the corresponding oxidation electrodes. If the electrode material is a nonmetal, as in hydrogen, oxygen and halogen electrodes, and in the so-called oxidation-reduction electrodes, an inert metal, frequently platinum, is employed to act as the conducting material. If the symbol Pt , for the inert metal, is written to the left, e.g., $\text{Pt}, \text{H}_2(g), \text{H}^+$; $\text{Pt}, \text{Cl}_2(g), \text{Cl}^-$; $\text{Pt}, \text{I}_2(s), \text{I}^-$; or $\text{Pt}, \text{Fe}^{+++}, \text{Fe}^{++}$, the convention is that oxidation is implied. On the other hand, if the symbol Pt is written to the right, as in $\text{H}^+, \text{H}_2(g), \text{Pt}$; $\text{Fe}^{+++}, \text{Fe}^{++}, \text{Pt}$, etc., the process is reduction and the sign of the potential is reversed. The order of writing the components of the solution, in the case of oxidation-reduction electrodes, e.g., $\text{Fe}^{+++}, \text{Fe}^{++}$ or $\text{Fe}^{++}, \text{Fe}^{+++}$, is immaterial.

When a cell is made by combining two reversible electrodes, then according to convention the left-hand electrode will exert its oxidation potential, while the right-hand one will exhibit its reduction potential, e.g., as in the Daniell cell which may be represented by



The E.M.F. of the complete cell is then equal to the algebraic *sum* of the potentials of the two electrodes, *one being an oxidation potential and the other a reduction potential*. Alternatively, the E.M.F. may be taken as equal to the *difference of the oxidation potentials* of the two electrode systems [cf. equation (80)]; the two methods of representation are, of course, identical. The application of these rules will be illustrated later.

The Hydrogen Scale.—There is no reliable method known for determining the absolute potential of a single electrode. The only sound procedure is to combine the electrode with a reference electrode of known potential difference and to measure the E.M.F. of the resulting cell. If the potential of the reference electrode is arbitrarily taken as zero, the

measured E.M.F. will be equal to the potential of the unknown electrode on this scale. The arbitrary zero of potential in general use is taken as that of a *reversible hydrogen electrode, with gas at 1 atm. pressure, in a solution of hydrogen ions of unit activity*. This particular electrode, i.e., $H_2 (1 \text{ atm.}) | H^+ (a_{H^+} = 1)$, is known as the **standard hydrogen electrode**, for the hydrogen and hydrogen ions are in their standard states. The convention is, therefore, to take the potential of the standard hydrogen electrode as zero; electrode potentials based on this arbitrary zero are said to refer to the **hydrogen scale**. The so-called "potentials on the hydrogen scale" are not really electrode potentials, for they are actually the E.M.F.'s of the cells obtained by combining the particular electrode with a standard hydrogen electrode; this fact must be borne in mind in connection with certain calculations based on these potentials (cf. p. 1010).

Subsidiary Reference Electrodes.—The use of the standard hydrogen electrode is not always convenient in practice, and so several subsidiary reference electrodes have been devised. The most common of these for general purposes is the **calomel electrode**, consisting of mercury in contact with solid mercurous chloride and a solution of potassium chloride saturated with the mercurous salt. Three different concentrations of potassium chloride have been employed, viz., 0.1 N, 1.0 N and a saturated solution. The potentials, on the hydrogen scale, of these three reference electrodes in the vicinity of 25° C. are as follows:

0.1 N KCl	$Hg_2Cl_2(s) Hg$	0.3338 — 0.00007 ($t - 25$) volt
1.0 N KCl	$Hg_2Cl_2(s) Hg$	0.2800 — 0.00024 ($t - 25$)
Saturated KCl	$Hg_2Cl_2(s) Hg$	0.2115 — 0.00076 ($t - 25$).

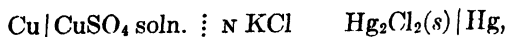
Because the calomel electrodes are employed as reference electrodes, for the determination of the oxidation potentials of other electrodes, they have been deliberately represented here as reduction electrodes. The electrode with 0.1 N potassium chloride is preferred for accurate work, as it has the lowest temperature coefficient, but the electrode with saturated potassium chloride solution is very convenient for many purposes and is often used. Various types of vessels have been designed for the purpose of setting up calomel electrodes, some of these are depicted in Fig. 238. The object of the special forms is generally to prevent diffusion of extraneous electrolytes into the potassium chloride.

In order to obtain the potential of any given electrode on the hydrogen scale, it is combined with the reference electrode and the E.M.F. of the resulting cell is measured; knowing the potential of the latter electrode that of the former can be derived. For example, combination of a Zn, $ZnSO_4$ soln. electrode with a calomel electrode containing N potassium chloride gives the cell



with an E.M.F. of 1.083 volt at 25°. Since this represents the sum of the oxidation potential of the zinc electrode and the reduction potential,

i.e., $+0.280$, of the calomel electrode, the former potential is $1.083 - 0.280 = 0.803$ volt. For a Cu, CuSO_4 soln. electrode the corresponding cell is



whose E.M.F. is -0.018 volt. The oxidation potential of the copper electrode is thus $-0.018 - 0.280 = -0.298$ volt.

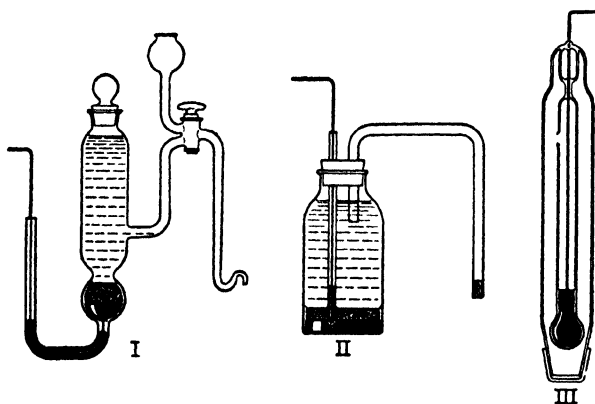
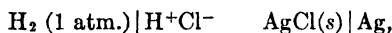


FIG. 238. Forms of calomel electrode

In recent years the Ag, AgCl electrode has been employed as a reference electrode, especially in connection with accurate measurements in chloride solutions, thus avoiding liquid junction potentials (p. 943). In this case it is the standard electrode potential which is particularly useful, and it is determined in the following manner. The E.M.F. of the cell



where the activities of the hydrogen and chloride ions in the hydrochloric acid solution have arbitrary values, is given by (79) as

$$E = E^0 - \frac{2RT}{F} \ln a, \quad (90)$$

where a is the mean activity of the hydrochloric acid. Upon adding $(2RT/F) \ln m$ to each side, where m is the molality of the acid solution in the given cell of E.M.F. equal to E , the result is

$$\left(E + \frac{2RT}{F} \ln m \right) - E^0 = - \frac{2RT}{F} \ln \frac{a}{m} = - \frac{2RT}{F} \ln \gamma, \quad (91)$$

where γ is the mean activity coefficient of the hydrogen and chloride ions in the solution. If the experimental quantity $E + (2RT/F) \ln m$ is plotted against a function of the molality, generally \sqrt{m} , as abscissa, and the curve extrapolated to $m = 0$, i.e., to infinite dilution, the ordinate gives E^0 , for at

infinite dilution γ is unity and the right-hand side of (91) is zero. The most accurate data, extrapolated by special methods, give E° as $+0.2224$ volt at 25°C. ; this is the E.M.F. of a hydrogen-silver chloride cell with hydrochloric acid at a mean activity of unity. For such a solution $a_{\text{H}^+} \times a_{\text{Cl}^-} = 1$, and so the E.M.F. of the cell



must be $+0.2224$ volt. By convention the left-hand potential is zero, and so the standard (reduction) potential of the silver-silver chloride electrode is $+0.2224$ volt at 25°C. The corresponding oxidation potential of the electrode



is -0.2224 volt. The silver-chloride electrode is readily prepared by depositing silver electrolytically on a platinum wire, and then coating it with finely divided silver by the decomposition of silver oxide. The silvered wire is now made an anode in a dilute solution of sodium chloride, using a very low current, when an adherent coating of silver chloride is formed. Although the electrode is highly reproducible, it must not be exposed to light, for obvious reasons.

Reproducible Liquid Junctions.—The combination of an electrode of unknown potential with a reference electrode, for the purpose of evaluating the former, often involves contact between two solutions of different electrolytes. In this way a liquid junction potential is introduced which must be either eliminated or for which allowance must be made. M. Planck (1890) and P. Henderson (1907–08) derived equations for the potential between any two solutions; the former postulated a sharp boundary, but the latter assumed the boundary to consist of a series of mixtures of the electrolytes in all proportions. The manner of setting up the liquid junction is, therefore, of importance. For accurate work the **flowing junction**, devised by A. B. Lamb and A. T. Larson (1920), and since modified in various ways, is the most reproducible. In its original form an upward current of the heavier electrolyte met a downward flow of the other at a point where a horizontal outlet tube joined the vertical tube through which the liquids entered; the liquids flowed at the same rate and a sharp boundary was maintained. The flowing junction probably approximates in behavior to the “continuous mixture” boundary postulated by Henderson, and his equation may be applied. For two binary electrolytes at concentrations c and c' moles per liter, respectively, it is

$$E_l = \frac{RT}{F} \cdot \frac{c(u_+ - u_-) - c'(u'_+ - u'_-)}{c(u_{+z_+} + u_{-z_-}) - c'(u'_{+z'_+} + u'_{-z'_-})} \ln \frac{c(u_{+z_+} + u_{-z_-})}{c'(u'_{+z'_+} + u'_{-z'_-})}, \quad (92)$$

where u_+ and u_- are the mobilities of positive and negative ions, and z_+ and z_- their valences. If for two uni-univalent electrolytes having an ion, e.g., the cation, in common, the concentrations are the same, i.e.,

$u_+ = u'_+$, $c = c'$, and all the z terms are unity, then (92) becomes

$$E_i = \frac{RT}{F} \ln \frac{u_+ + u'_-}{u_+ + u_-} = \frac{RT}{F} \ln \frac{\Lambda'}{\Lambda}. \quad (93)$$

This is the equation of G. N. Lewis and L. W. Sargent (1909), which can also be derived from the Planck equation for the liquid junction potential.

For approximate work the junction may be made in a parchment or collodion membrane, in a gel or in a capillary tube, so as to minimize the effects of diffusion; alternatively, the end of one of the tubes containing the electrolyte may be plugged with a tight roll of filter paper or absorbent cotton. A simple and useful form of liquid junction is set up by allowing the liquids to come into contact in the narrow space between a ground-glass plug and its seating; the resistance of the cell is high, but this is no serious disadvantage with modern methods of measuring E.M.F.

For most laboratory purposes, where high precision is not required, the practice is to eliminate the potential, or to reduce it to a negligible amount; this is achieved by the use of a **salt bridge**, consisting of a concentrated solution of either potassium chloride or of ammonium nitrate, the former being preferred. The efficacy of these substances is probably connected with the fact that in each case the anion and cation have almost the same velocity. Since the concentration of the salt forming the bridge is much larger than that of other electrolytes, it is responsible for the transfer of almost the whole of the current across the liquid junction; it will be evident, therefore, that if the transference numbers of the ions are about 0.5 the potential will be small (cf. p. 932). The use of the calomel electrode with saturated potassium chloride solution, in conjunction with a salt bridge of the same electrolyte, eliminates one junction entirely. Whenever it is assumed that the liquid junction potential is annulled two vertical lines are employed in the representation of the cell.²⁰

Standard Electrode Potentials.—The important property of an electrode is clearly its standard potential, which has a definite value at a given temperature; * from it the potential of an electrode with the constituents at arbitrary activities may, in principle, be calculated by the equations given above, e.g., (83) to (86). Although there is no fundamental difference between electrodes of different types, as already seen, it is nevertheless convenient to consider three categories separately, viz., electrodes reversible with respect to positive ions, with respect to negative ions, and oxidation-reduction electrodes.

The basis of the method for determining the standard potential of a metal in contact with a solution of its own (positive) ions is to measure the potential E of an electrode containing the ions at an arbitrary activity, and then to derive the standard potential E^0 by the relationship

$$E_+ = E^0 - \frac{RT}{zF} \quad (94)$$

* The external pressure is assumed to be constant at 1 atm.

In applying this equation there is, however, the fundamental difficulty that it is not possible, without making assumptions, to evaluate the activities of individual ionic species.* In some instances, e.g., the silver-silver chloride electrode already considered, the difficulty can be circumvented, but for most purposes the mean activity of the ions may be used. For approximate requirements the activity (a) may be replaced by the concentration (c) of the ions, assuming complete dissociation; the electrode potential equation may then be written in the approximate, but convenient, form

$$E_+ = E^0 - \frac{RT}{zF} \ln c_+. \quad (95)$$

On this basis the standard potential is that in a solution containing 1 g. ion per liter of the reversible ion.

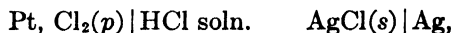
The standard electrode potentials, of varying degrees of accuracy, for a number of metals are quoted in Table 156 for 25° c. The values are

TABLE 156. STANDARD (OXIDATION) POTENTIALS OF METALS AT 25° C.

$H_2, H^+ = 0$					
Li, Li^+	+3.024	Zn, Zn^{++}	+0.761	Pb, Pb^{++}	+0.126
K, K^+	+2.924	Fe, Fe^{++}	+0.441	Bi, Bi^{+++}	-0.226
Na, Na^+	+2.715	Cd, Cd^{++}	+0.402	Cu, Cu^{++}	-0.340
Mg, Mg^{++}	+2.34	Co, Co^{++}	+0.28	Ag, Ag^+	-0.799
Al, Al^{+++}	+1.67	Ni, Ni^{++}	+0.24	Hg, Hg_2^{++}	-0.799
Mn, Mn^{++}	+1.1	Sn, Sn^{++}	+0.140	Au, Au^{+++}	-1.30

the oxidation potentials in each case, so that they correspond to the process $M \rightarrow M^+ + \text{electrons}$. There are many uses of the standard electrode potentials, reference to some of which will be made later; in the meantime the most obvious application may be noted. Since E^0 for a given metal is available, then from a knowledge of the potential in any solution of its ions the approximate activity or concentration may be evaluated by (94) or (95).

Anion Electrodes.—The most important anion electrodes are the halogen and oxygen electrodes. In the cell



with chlorine at a relatively low pressure p , the cell reaction is



and since the silver chloride and silver are in the solid state, i.e., at unit

* Individual ion activities could only be determined if liquid junction potentials might be calculated exactly or eliminated entirely; the calculation involves a knowledge of ionic activities [cf. equation (66)], and so must be ruled out, and the complete elimination is uncertain.

activity, the E.M.F. is given by (76) as

$$E = E^0 - \frac{RT}{2F} \ln p. \quad (96)$$

By measuring E , the E.M.F. of the cell with the chlorine gas at the known pressure p , the standard E.M.F. of the cell can be determined from (96); in this way E^0 is found to be -1.136 volt at 25°C . As seen above, the standard E.M.F. of the cell is equal to the difference in the standard oxidation potentials of the two electrodes; that of the silver chloride electrode is known to be -0.222 volt (p. 943), and so the value for the chlorine gas electrode must be $-1.136 - 0.222$, i.e., -1.358 volt. Somewhat similar methods have been used for the bromine and iodine electrodes; in the former case the standard state may be chosen as either the vapor at 1 atm. pressure or the pure liquid, while for the latter the solid is taken as the standard state. Corrections must be applied for attack of the free halogen, e.g., bromine, on water, and for the formation of perhalide ions, e.g., with iodine. The standard potential of the fluorine electrode is calculated by indirect methods (see p. 1013).

Since the oxygen electrode does not behave reversibly, its standard potential, with respect to OH^- ions, is obtained in the following manner. The reaction taking place in the cell



is represented by



for the passage of two faradays. From a variety of data, the best probably being those based on partition functions (p. 879), it is known that the standard free energy change for this reaction, that is, with both gases at 1 atm. pressure combining to form pure liquid water, is $-56,700$ cal. at 25°C . This will be equal to $-2FE^0$, where E^0 is the standard E.M.F. of the hydrogen-oxygen cell depicted above. Since 1 volt-coulomb is equivalent to 0.239 cal., it follows that E^0 is equal to $56,700/2 \times 96,500 \times 0.239 = 1.229$ volts at 25°C . If the aqueous solution contained hydrogen ions at unit activity, the potential of the left-hand electrode would be zero, and hence it follows that the (reduction) potential of the $\text{H}^+(\alpha_{\text{H}^+} = 1)$, O_2 (1 atm.), Pt electrode is 1.229 volt. It will be seen on p. 992 that in an aqueous solution, provided the activity of the water is unity, the product $\alpha_{\text{H}^+}\alpha_{\text{OH}^-}$ is 1.008×10^{-14} at 25° , and so this must be the activity of the hydroxyl ions in the solution under consideration. Utilizing the appropriate form of (85), viz.,

$$E = E_{\text{O}_2}^0 + \frac{RT}{F} \ln \alpha_{\text{OH}^-} = E_{\text{O}_2}^0 + 0.05915 \log \alpha_{\text{OH}^-} \quad (97)$$

at 25° , and taking E as 1.229 and α_{OH^-} as 1.008×10^{-14} , it is readily found that E^0 is 0.401 for the oxygen electrode. The standard oxidation potential for the oxygen electrode $\text{Pt}, \text{O}_2(g), \text{OH}^-$, is thus -0.401 volt at 25°C .

The method of deriving the standard potential of the silver-silver chloride electrode, which is reversible with respect to chloride ions, was

described on p. 942; a similar procedure has been adopted for analogous halide electrodes. Some of the standard (oxidation) potentials for anion electrodes are recorded in Table 157 for 25° c.; the reaction in each case

TABLE 157. STANDARD (OXIDATION) POTENTIALS OF ANION ELECTRODES AT 25° C.

Pt, $\frac{1}{2}\text{F}_2$ (1 atm.), F^-	-2.85 volt	Pt, $\frac{1}{2}\text{O}_2$ (1 atm.), 2OH^-	-0.401 volt
Pt, $\frac{1}{2}\text{Cl}_2$ (1 atm.), Cl^-	-1.358	Hg, $\frac{1}{2}\text{Hg}_2\text{Cl}_2(\text{s})$, Cl^-	-0.2680
Pt, $\frac{1}{2}\text{Br}_2$ (1 atm.), Br^-	-1.087	Ag, $\text{AgCl}(\text{s})$, Cl^-	-0.2224
Pt, $\frac{1}{2}\text{Br}_2(\text{l})$, Br^-	-1.065	Ag, $\text{AgBr}(\text{s})$, Br^-	-0.0711
Pt, $\frac{1}{2}\text{I}_2(\text{s})$, I^-	-0.535	Ag, $\text{AgI}(\text{s})$, I^-	+0.1522

may be represented by $\text{A}^- \rightarrow \text{A} + \text{electrons}$, e.g., $\text{Cl}^- \rightarrow \frac{1}{2}\text{Cl}_2 + \epsilon$, $\text{Ag} + \text{Cl}^- \rightarrow \text{AgCl} + \epsilon$, and $2\text{OH}^- \rightarrow \frac{1}{2}\text{O}_2 + \text{H}_2\text{O} + 2\epsilon$.

Influence of Concentration and Valence on Electrode Potential.—To investigate the effect of concentration and valence on electrode potential, numerical values may be inserted in (86) for R and F ; the former is 8.313 int. joules deg.⁻¹ mole⁻¹, and the latter is 96,500 int. coulombs, and multiplying by 2.3026 to convert the logarithms, it is seen that

$$E_{\pm} = E^0 \mp \frac{2.3026 \times 8.313}{z_{\pm} \times 96,500} T \log a_i \quad (98)$$

$$= E^0 \mp \frac{0.0001984}{z_{\pm}} T \log a_i \approx E^0 \mp \frac{0.0002}{z_{\pm}} T \log a_i. \quad (99)$$

At 25° c., i.e., 298.16° K., (99) becomes

$$E_{\pm} = E^0 \mp \frac{0.05915}{z_{\pm}} \log a_i, \quad (100)$$

so that at ordinary temperatures a ten-fold change in the activity (or concentration) of an ion changes the potential by about 0.059/ z volts. In general, a change of concentration by a factor of 10^n alters the potential by $n \times 0.059/z$ volts at laboratory temperatures.

An application of the influence of concentration on potential is to determine the valence of the mercurous ions. A ten-fold change in concentration of mercurous nitrate alters the potential of mercury by 0.029 volt; since this is approximately equal to 0.059/2, it is evident that the valence z is two. Mercurous ions are thus to be represented as Hg_2^{2+} .

Mechanism and Significance of Electrode Potential.—The thermodynamic treatment permits an exact derivation of the expression for the E.M.F. of a reversible cell or the potential of an electrode, but it throws no light on the mechanism of the establishment of the potential difference. The first reasonable suggestions were made by W. Nernst (1889), who considered that all metals, and hydrogen, possess a property, called **solution pressure**, in virtue of which they tend to pass into solution as positive ions. In doing so they leave the metal negatively charged, and an electrical double layer (see Chapter XIV) is set up at the electrode surface. As a result of the attraction of the opposite charges the amount of metal that goes into solution is minute, and

the double layer, with a definite difference of potential, is established rapidly. According to Nernst the tendency for the metal to send ions into solution is opposed by the reverse tendency of the ions in solution, as a consequence of their osmotic pressure, to leave the solution and be deposited on the solid metal. Similar considerations would apply to nonmetals, such as oxygen and the halogens, whose solution pressure would be a measure of the tendency to send negative ions into solution. By the use of thermodynamic methods, based on these postulates, an equation of the form of (86), at one time commonly known as the **Nernst equation**, was derived. From the argument it appeared that the standard potential of a metal or nonmetal was determined by its hypothetical solution pressure. Because of the doubtful validity of the thermodynamic treatment and the uncertainty of the physical significance of the so-called solution "pressure," the views of Nernst, which played an important rôle in the development of electrochemistry, may now be regarded as obsolete.

At the present time the establishment of electrode potential is generally attributed to the rates of two opposing processes, viz., the passage of ions from the metal (or nonmetal) into solution and the reverse, i.e., discharge of the ions in the solution to form atoms of the metal (or nonmetal) on the electrode. If, with a metal electrode, the former process is the more rapid, cations will pass into solution more rapidly than they can return. As a result, free electrons will be left on the metal and positive ions will accumulate on the solution side of the electrode, thus building up an electrical double layer; the potential difference across this double layer then represents the electrode potential. However, the existence of this potential makes it more difficult for positive ions to leave the negatively charged metal and enter the solution, while the transfer of ions to the metal is facilitated. When equilibrium is established the two processes are occurring at the same rate, and the electrode exhibits its reversible potential.

If E is the actual potential difference across the double layer, formed by the electrons and positive ions at the metal-solution interface, it may be supposed that a fraction α facilitates the discharge of ions, while the remainder, $1 - \alpha$, hinders the passage of ions into solution. The actual value of α is immaterial, since it cancels out at a later stage. In its transfer across the double layer, therefore, the free energy of the discharging ion is increased by αzFE , where z is its valence, whereas the free energy of the atom passing into solution is diminished by $(1 - \alpha)zFE$. The result of these free energy changes is that in the presence of the double layer potential E , the rates of the reactions may be written as follows:

$$\text{Rate of discharge of ions from solution} = kc_+e^{\alpha zFE/RT}$$

$$\text{Rate of passage of ions into solution} = k'e^{-(1-\alpha)zFE/RT},$$

where k and k' are the rate constants. The former of these rates is seen to depend on c_+ , the concentration of the ions in the solution, but the latter does not involve a concentration factor since the "active mass" of the solid metal is constant. At the equilibrium, i.e., reversible, potential the rates of the two processes must be equal; hence

$$kc_+e^{\alpha zFE/RT} = k'e^{-(1-\alpha)zFE/RT},$$

$$\therefore E = \frac{RT}{zF} \ln \frac{k}{k'} - \frac{RT}{zF} \ln c_+. \quad (101)$$

Comparison of (101) with (84) shows that the former is of the correct form, at least for an ideal solution where the activity may be taken as equal to the concentration. The standard potential of the electrode is seen to be determined by k and k' , the rate constants of the two processes occurring at the electrode.

A closer insight into the significance of the standard potential may be obtained from another point of view. The absolute standard single potential of a metal is a measure of the standard free energy change of the reaction



that is, the formation of solvated ions, at unit activity, in the solution. In the previous discussion the solvation of the ions has been ignored, but it must not be forgotten. The process depicted above may be regarded as taking place in a series of stages, for each of which the free energy change may be related to a definite physical property.

I. An atom of the metal is vaporized; the free energy increase is equal to the energy (or heat) of sublimation, S .

II. The atom of the vapor is ionized to form gaseous metal ions M^{z+} , and z electrons. The energy increase is $I\epsilon$ per atom, where I is the ionization potential of the metal and ϵ represents the electronic charge; for 1 g. atom this becomes $IN\epsilon$, or IF .

III. The gaseous metal ion is dissolved in the solvent so as to form solvated ions in solution at unit activity; the increase of energy is represented by $-W_{M^+}$ (cf. p. 412).

IV. The z electrons produced in stage II are returned to the metal. The energy change is $-ze\phi$ per atom or $-zF\phi$ per g. atom, where ϕ is the electronic work function of the metal (p. 33).

The total energy change in these four stages may be identified (approximately) with ΔF^0 and hence with $-zFE^0$, where E^0 is now the absolute value of the single standard electrode potential; thus,

$$-zFE^0 = S + IF - W_{M^+} - zF\phi. \quad (102)$$

The standard potential is thus determined by the heat of sublimation, the ionization potential and the work function of the metal, and also by the heat of solvation of the ions.²¹

Oxidation-Reduction Electrodes.*—The earliest studies of the potentials of oxidation-reduction systems were made by W. D. Bancroft (1892), but many of his electrodes were not reversible in the thermodynamic sense. A reversible oxidation-reduction electrode is one in which both oxidized and reduced stages are present, each being convertible into the other by an infinitesimally small change of potential from the equilibrium value. Examples of reversible systems are ferric-ferrous, stannic-stannous, ferricyanide-ferrocyanide, and quinone-hydroquinone. In each of these solutions a platinum electrode will acquire a definite potential dependent on the *ratio* of the oxidized to reduced form, as required by (83) or (89). The particular forms taken by the general equation in some special cases have already been given, in (87) and (88), and others will be considered shortly.

* Sometimes referred to as "redox" electrodes.

In principle the method used for evaluating the standard potential of an oxidation-reduction system is based on (83) or (89); the potential E acquired by an inert, e.g., platinum, electrode when inserted in the solution containing the oxidized and reduced states at known activities, is measured, and E^0 may then be calculated from the appropriate form of (83). In practice some difficulty is encountered because of the marked difference between activities and concentrations of high valence ions, e.g., Fe^{+++} , Sn^{++++} , $\text{Fe}(\text{CN})_6^{-----}$, such as are encountered in oxidation-reduction systems. By the use of suitable extrapolation and other methods this source of error has been overcome in many instances; some of the results obtained are recorded in Table 158 for 25° c. As in the two pre-

TABLE 158. STANDARD (OXIDATION) POTENTIALS OF OXIDATION-REDUCTION SYSTEMS

Pt, Co^{++} , Co^{+++}	-1.82	Pt, Fe^{++} , Fe^{+++}	-0.771
Pt, Pb^{++} , Pb^{++++}	-1.75	Pt, MnO_4^{--} , MnO_4^-	-0.54
Pt, Ce^{+++} , Ce^{++++}	-1.61	Pt, $\text{Fe}(\text{CN})_6^{-----}$, $\text{Fe}(\text{CN})_6^{----}$	-0.356
Pt, Mn^{++} , $\text{MnO}_4^- + 8\text{H}^+$	-1.52	Pt, Cu^+ , Cu^{++}	-0.16
Pt, Ti^+ , Ti^{+++}	-1.25	Pt, Sn^{++} , Sn^{++++}	-0.15
Pt, $\frac{1}{2}\text{I}_2$, $\text{IO}_3^- + 6\text{H}^+$	-1.20	Pt, Ti^{+++} , Ti^{++++}	-0.06
Pt, $\frac{1}{2}\text{Hg}_2^{++}$, Hg^{++}	-0.91	Pt, Cr^{++} , Cr^{+++}	+0.41

ceding tables, the values given are "oxidation" potentials; this is indicated by the symbol Pt, for the inert electrode, which precedes the formula for the system in each case.

Standard E.M.F.'s and Equilibrium Constants.—The equation (77), giving the relationship between the standard E.M.F. of a cell and the equilibrium constant of the cell reaction, may be written in the form of (100), i.e.,

$$E^0 = \frac{0.05915}{n} \log K \quad (103)$$

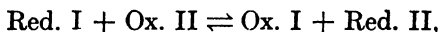
at 25° c. The value of E^0 can be derived from data similar to those in Tables 156, 157 and 158, and hence the equilibrium constant K can be evaluated. For example, the reaction $\text{Zn} + \text{Cu}^{++} \rightleftharpoons \text{Zn}^{++} + \text{Cu}$ occurs in the Daniell cell, $\text{Zn} | \text{ZnSO}_4 \text{ soln.} || \text{CuSO}_4 \text{ soln.} | \text{Cu}$; by Table 156, E^0 is +0.761 for the zinc electrode and -0.340 for the copper electrode, and so E^0 for the cell is +0.761 - (-0.340), i.e., 1.101 volt. Since n is 2 for the reaction under consideration, it follows from (103) that K , equal in this case to $a_{\text{Zn}^{++}}/a_{\text{Cu}^{++}}$, or approximately $c_{\text{Zn}^{++}}/c_{\text{Cu}^{++}}$, is about 10^{37} at ordinary temperatures.* This result means that if zinc is placed in a solution of copper sulfate, it will continue to dissolve until the ratio of the concentration of zinc to copper ions in the solution is 10^{37} ; that is, zinc will replace copper completely, within the limits of analytical error. The same method may be used to determine the extent to which any

* The same result is obtained from the fact that at equilibrium the E.M.F. of the cell is zero, and the two electrodes must have the same potential; thus, at 25° c., $-0.340 - (0.059/2) \log a_{\text{Cu}^{++}} = +0.761 - (0.059/2) \log a_{\text{Zn}^{++}}$, where the activities are the values at equilibrium.

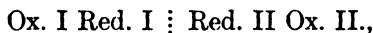
metal will displace another; the results will be generalized below so as to include all types of oxidation and reduction.

If the equilibrium constant can be determined experimentally, E^0 may be evaluated and hence the standard potential of one electrode may be obtained if the other is known. For example, when finely divided lead and tin were shaken with a solution containing lead and tin perchlorates, it was found that at equilibrium $c_{\text{Sn}^{++}}/c_{\text{Pb}^{++}}$, which may be assumed to be equal to K , was 2.98 at 25° C.; substitution in (103), with n equal to 2, gives $E^0 = 0.014$. Since the solution at equilibrium contains more tin than lead ions, the tin electrode must have the more positive potential. The standard potential, Pb, Pb^{++} is +0.126, and so Sn, Sn^{++} must be +0.140 volt; the value recorded in Table 156 was obtained in this manner.

The results described above may be made quite general, so as to include systems of all types which yield reversible electrodes. As seen on p. 939, any electrode oxidation reaction may be written as: reduced state \rightarrow oxidized state + electrons, and the reduction process is the reverse, i.e., oxidized state + electrons \rightarrow reduced state. The general reaction between two systems, designated I and II, may be written as



assuming, for simplicity, the same number of electrons to be involved in each process. This reaction will take place in the cell



for which the standard E.M.F. is given by $E^0_{\text{I}} - E^0_{\text{II}}$. It follows, therefore, from (77) that

$$E^0_{\text{I}} - E^0_{\text{II}} = \frac{RT}{nF} \ln \left(\frac{\text{Ox. I} \times \text{Red. II}}{\text{Red. I} \times \text{Ox. II}} \right)_e, \quad (104)$$

where the subscript e implies that the activities in the parentheses are the equilibrium values. It is seen from (104) that provided E^0_{I} is larger algebraically than E^0_{II} the equilibrium of the general reaction will lie to the right, that is to say, system I will be oxidized by II, while II is reduced by I. The greater the difference in the standard potentials the greater will be the equilibrium constant, and the more completely will the reaction take place. This conclusion is applicable to any pair of systems for which data are given in Tables 156, 157 and 158, irrespective of their type. In general, for any process to be sufficiently complete for quantitative (volumetric) analytical purposes the standard potentials should differ by at least 0.3 volt; this figure is not exact, as it depends on the numbers of electrons concerned in the two systems, but it may be taken as a rough guide.

Organic Oxidation-Reduction Systems.—A number of organic oxidation-reduction systems behave in a reversible manner and the electrode potentials satisfy the requirements of (83); the most familiar kind in-

volves a quinone and a hydroquinone, e.g., quinone itself, viz.,



This system was first studied by F. Haber and R. Russ (1904), but the same type of equilibrium, in which two hydrogen ions and two electrons take part, occurs in other instances. In the general case



the potential acquired by an unattackable electrode, e.g., platinum, when placed in a solution containing Q, H_2Q and hydrogen ions, should be given by

$$\begin{aligned} E &= E^0 - \frac{RT}{2F} \ln \frac{a_{\text{Q}} a_{\text{H}^+}^2}{a_{\text{H}_2\text{Q}}} \\ &= E^0 - \frac{RT}{2F} \ln \frac{a_{\text{Q}}}{a_{\text{H}_2\text{Q}}} - \frac{RT}{F} \ln a_{\text{H}^+}. \end{aligned} \quad (105)$$

The potential of the Q- H_2Q systems, as of most systems involving non-ionized organic compounds, is dependent on the hydrogen ion activity of the solution; this fact is utilized for the purpose of determining the latter quantity (p. 995). It should be noted that (105) holds only if the substance H_2Q does not ionize as an acid to yield hydrogen ions; in solutions of low hydrogen ion concentration, i.e., in alkaline media, the H_2Q compounds, e.g., hydroquinone, are ionized, and it can be shown that if this is appreciable the potential of the system becomes independent of the hydrogen ion activity.

The standard potential of the system under discussion is that at which Q, H_2Q and H^+ are at unit activity, but for many purposes it is convenient to define a secondary standard potential applicable at a given hydrogen ion concentration; the value generally chosen is 10^{-7} g. ion per liter, i.e., the neutral point, and the symbol E^0' is employed. This scheme has been adopted particularly for the oxidation-reduction indicators used in biological work, for the medium is generally approximately neutral.

Oxidation-Reduction Indicators.—If the forms Q and H_2Q of an organic compound have very different colors, as is often the case, mixtures in various proportions will have different shades of color. If the activity ratio $a_{\text{Q}}/a_{\text{H}_2\text{Q}}$ is approximately the same as the concentration ratio, then it is apparent from (105), that provided the hydrogen ion activity is kept constant, each color shade corresponds to a definite potential of the oxidation-reduction system. A substance of the type under consideration may, therefore, be used as an **oxidation-reduction indicator**. If a small amount is placed in another oxidation-reduction system which is to be studied, the two systems interact until equilibrium is attained when they both have the same potential. Since the system under examination is present in large excess the amounts of oxidized and reduced forms are virtually unaltered, and the indicator system adjusts itself so as to acquire

the same potential as that of the system in which it is placed. The color of the indicator then gives the potential of the latter, provided either it has been standardized in solutions of known potential, or its $E^{\circ'}$ value at the hydrogen ion concentration of the experimental system is known; in the latter case the ratio $c_Q/c_{H_2Q} \approx a_Q/a_{H_2Q}$ is estimated from the depth of color and (105) used to calculate E . Owing to the limitation of the eye, or even of mechanical devices, for the detection of color changes beyond certain ratios, a given indicator can only be used for the determination of potentials lying about 0.03 volt either side of the $E^{\circ'}$ value under the given conditions. It is necessary, therefore, to have a series of such indicators, some of which are given in Table 159; the subsidiary standard

TABLE 159. STANDARD POTENTIALS IN NEUTRAL SOLUTION OF SOME OXIDATION-REDUCTION INDICATORS

Indicator	$E^{\circ'}$	Indicator	$E^{\circ'}$
Phenol- <i>m</i> -sulfonate indo-2:6-dibromophenol	-0.273	Ethyl Capri blue	+0.072
2:6-Dichlorophenol indophenol	-0.217	Indigo disulfonate	+0.125
Toluylene blue	-0.115	Cresyl violet	+0.173
Cresyl blue	-0.047	Phenosafranine	+0.252
Methylene blue	-0.011	Rosinduline scarlet	+0.296
Indigo tetrasulfonate	+0.046	Sulfonated rosindone	+0.380
		Methyl viologen	+0.445

$E^{\circ'}$ refers to a neutral solution. The use of oxidation-reduction indicators for the purpose of volumetric analysis will be considered later (p. 1009).²²

E.M.F.'s in Nonaqueous Media.—The thermodynamic arguments whereby the equations for E.M.F. and electrode potentials were deduced should hold irrespective of the nature of the solvent, and experiments show that the fundamental behavior of electrodes is the same in aqueous as in nonaqueous media. At one time doubt was cast on the applicability of the equation for the E.M.F. of a simple concentration cell in the latter type of solvent, but the observed deviations were undoubtedly due to the considerable difference between the activity ratio, which should be used theoretically, and the concentration ratio, which was generally employed in practice. For the measurement of single electrode potentials it is necessary to have a standard, as described on p. 941; the earlier workers compared potentials in nonaqueous solutions with a reference electrode containing an aqueous electrolyte, but in this way large and uncertain liquid junction potentials were introduced. In more recent work it has become the practice to treat each solvent as an independent system, and to define the arbitrary zero of potential, in terms of the hydrogen electrode, in each individual solvent. The arbitrary zero in water may be completely different from the arbitrary zero of potential in another liquid, e.g., alcohol, and there is at present no satisfactory method for comparing them. The potentials measured in each medium are comparable with one another, but a *quantitative* comparison of the results for two solvents has no significance. Standard electrode potentials have been determined in methyl

and ethyl alcohols and in ammonia, and it is of interest to record that when the values are arranged in order, the series is very much the same in each case as in water; the alkali metals have the most positive potentials and the noble metals, e.g., silver and mercury, are the most negative. The halogens also have relatively high negative potentials. According to the arguments on p. 949 the standard potential depends on the energy of sublimation, the ionization potential, the electronic work function, and the energy of solvation of the ions. It is only the last of these quantities which varies with the nature of the solvent, and if it is relatively small compared with the others, the potential series of the elements in different solvents should not vary greatly.

ACTIVITY COEFFICIENTS

Application of Law of Equilibrium.—In a solution of any electrolyte there should be an equilibrium between free ions and undissociated molecules; the latter may consist of true nonionized molecules or of ion-pairs. The equilibrium can then be written in the form



where M^+ and A^- represent the free ions and MA is the undissociated portion of the electrolyte, including both nonionized molecules and ion-pairs. According to the law of equilibrium,

$$K = \frac{a_{M^+} \times a_{A^-}}{a_{MA}}, \quad (106)$$

where the a 's are the activities of the indicated species; the equilibrium constant K is called the **dissociation constant** * of the electrolyte. Writing the activities in (106) as the product of the concentration c , in g. ions or moles per liter, and the activity coefficient f (see p. 685), it becomes

$$K = \frac{c_{M^+} \times c_{A^-}}{c_{MA}} \cdot \frac{f_{M^+} \times f_{A^-}}{f_{MA}}. \quad (107)$$

Further, if α is the degree of dissociation of the electrolyte, i.e., the fraction of the electrolyte in the form of free ions, and c is its total concentration in moles per liter, c_{M^+} and c_{A^-} are each equal to αc , while c_{MA} is equal to $c(1 - \alpha)$; hence, (107) becomes

$$K = \frac{\alpha^2 c}{1 - \alpha} \cdot \frac{f_{M^+} \times f_{A^-}}{f_{MA}}. \quad (108)$$

If the solution is sufficiently dilute, the activity coefficients are approxi-

* The term "ionization constant" is frequently employed, but since the equilibrium is between free ions and *undissociated* molecules, the expression "dissociation constant" is preferred.

mately unity, and so (108) reduces, under these conditions, to

$$k = \frac{\alpha^2 c}{1 - \alpha}, \quad (109)$$

which is the expression of the **dilution law**, as first derived by W. Ostwald (1888).²³ It should be noted that k in (109) is not really a constant, because of the neglect of the activity coefficients, and so the symbol k is used in place of K , which is a true (thermodynamic) constant.

In the earliest attempts to verify (109) the value of α was taken as equal to Λ/Λ_0 ; that is, the conductance ratio was assumed to be identical with the degree of dissociation. This is known to be approximately true (p. 903), and the application to solutions of acetic acid is exemplified by the data in Table 160 (D. A. MacInnes and T. Shedlovsky, 1932); Λ_0 is taken as 390.7 ohms⁻¹ cm.² at 25° C. The values of k , in the third column, show approximate constancy. The same type of behavior is found for many weak electrolytes, such as weak acids and bases, and for many years the conductance method was regarded as giving satisfactory dissociation constants for such electrolytes. Many of the data recorded in the older literature were obtained in the manner described. A closer examination of the results shows, however, that with increasing concentration, and particularly if the electrolyte is not especially weak, the so-called dissociation constant exhibits a definite drift. The drift is partly due to the fact that the conductance ratio is not equal to the degree of dissociation; the latter is given more accurately by Λ/Λ' , as described on p. 902. Utilizing the results obtained in this manner, (109) gives the values k_{corr} in Table 160. In the very dilute solutions there may be an

TABLE 160. DISSOCIATION CONSTANT OF ACETIC ACID AT 25° C.

$c \times 10^3$	Λ	$k \times 10^5$	$k_{\text{corr}} \times 10^5$	$K \times 10^5$
0.028014	210.38	1.760	1.768	1.752
0.11135	127.75	1.769	1.779	1.754
0.2184	96.49	1.769	1.781	1.751
1.02831	48.15	1.781	1.797	1.751
2.4140	32.22	1.789	1.809	1.750
5.9115	20.96	1.799	1.823	1.749
9.8421	16.37	1.803	1.834	1.747
20.000	11.57	1.806	1.840	1.737
50.000	7.358	1.808	1.849	1.720

improvement in the constancy, but it does not persist, and so it appears that another factor is responsible for the steady variation in both "constants." This is undoubtedly the activity coefficient factor which was assumed to be unity in deriving (109) from (108). By making the necessary corrections, with the aid of the Debye-Hückel equations (cf. p. 970), which is particularly applicable in dilute solutions, the results in the last column of Table 160 were obtained. It is seen that K is almost exactly constant in all the less concentrated solutions; the deviations from con-

stancy at the higher concentrations are due to the failure of the simple Debye-Hückel activity equation under these conditions.

The Ionic Strength.—In order to represent the variation of activity coefficients with concentration, G. N. Lewis and M. Randall (1921) introduced the concept of the **ionic strength**, which is a measure of the intensity of the electrical field due to the ions in the solution. It is given the symbol μ and is defined as *half the sum of the terms obtained by multiplying the molality (or concentration) of each ionic species present in the solution by the square of its valence*; thus,

$$\mu = \frac{1}{2} \sum c_i z_i^2, \quad (110)$$

where c_i is the *actual* concentration (or molality) of each ion and z_i is its valence. It was pointed out by Lewis and Randall that, in dilute solutions, the activity coefficient of a given electrolyte is approximately the same in all solutions of the same ionic strength. The particular ionic strength may be due to the presence of added electrolytes, but their nature does not affect the activity coefficient of the electrolyte under consideration. If the ionic strength of the medium is kept constant, e.g., by the addition of neutral salts, the activity coefficients of the ions of a weak electrolyte should remain constant. Under these conditions the quantity k as given by (109) will be constant, although not equal to the true constant K ; the value of k , like the activity coefficients, depends on the actual ionic strength of the medium.

The Debye-Hückel Theory.—Until 1923 the activity coefficient was regarded as an empirical quantity, but in that year P. Debye and E. Hückel, following on the earlier work of S. R. Milner (1912), made the most significant advance in electrochemistry since Arrhenius enunciated the theory of electrolytic dissociation. By taking into account the inter-ionic forces, as described in relation to conductance on p. 903, it was found possible to calculate the ratio of the activity to the concentration of an ion in dilute solution. The basic principle is that as a consequence of the existence of an oppositely charged ionic atmosphere surrounding each ion, dilution of a solution involves work against the electrostatic attraction, in addition to the free energy of dilution for an ideal solution which is determined only by the decrease of ionic concentration. The difference between actual and ideal free energy changes is a measure of the activity coefficient of the particular ion in the given solution.

The electrical potential at any point is defined as the work done in bringing a unit charge from infinity to the particular point. Let ψ be the potential at a given point in the vicinity of a positive ion; then the work done in bringing a positive ion of valence z_+ , that is, carrying a charge $+z_+e$, where e is the electronic charge, to that point is $z_+e\psi$; similarly the work required for a negative ion is $-z_-e\psi$, where z_+ and z_- are *numerical* values only, and do not include the sign. If the concentrations of ions at a considerable distance from the given ion, where ψ may be taken as zero, are n_+^0 and n_-^0 , per unit volume, then by the

distribution law for particles in a field of varying potential (p. 272), the concentrations of positive and negative ions, i.e., n_+ and n_- , at the point under consideration, are given by

$$n_+ = n_+^0 e^{-(z_+ e \psi / kT)} \quad \text{and} \quad n_- = n_-^0 e^{-(-z_- e \psi / kT)}, \quad (111)$$

where k is the Boltzmann constant. Since z_+ and z_- are numerical only, it is clear that n_- is greater than n_+ , and so there are on the average more negative than positive ions in the vicinity of any positive ion; the reverse will, of course, be true for a negative ion. It is seen, therefore, as described on p. 903, that every ion is surrounded by an oppositely charged ionic atmosphere; that is to say, ions of opposite sign predominate in the ionic atmosphere.

The density of electricity ρ_e at any point is equal to the excess positive or negative electricity per unit volume at that point; that is,

$$\rho_e = n_+ z_+ e - n_- z_- e = n_+^0 z_+ e e^{-z_+ e \psi / kT} - n_-^0 z_- e e^{-z_- e \psi / kT} \quad (112)$$

In the simple case of a uni-univalent electrolyte, $z_+ = z_- = 1$, and n_+^0 and n_-^0 are equal; replacing the latter by n , the number per cc. of ions of either kind in the bulk of the solution, it follows from (112) that

$$\rho_e = n e (e^{-e \psi / kT} - e^{e \psi / kT}). \quad (113)$$

If the assumption is now made that $e \psi / kT$ is small in comparison with unity, then by writing out the exponential series and neglecting all terms beyond the first, (113) becomes

$$\rho_e = -(\epsilon^2 \psi / kT) 2n. \quad (114)$$

In the general case, when z_+ and z_- are not necessarily unity, and the solution may contain several different kinds of ions, the appropriate form of (112) is

$$\rho_e = e \sum n_i z_i e^{-z_i e \psi / kT}, \quad (115)$$

where n_i and z_i represent the number per cc. and valence, respectively, of each kind of ion in turn. If, as before, $z_i e \psi / kT$ is small compared with unity,

$$\rho_e = -(\epsilon^2 \psi / kT) \sum n_i z_i^2. \quad (116)$$

The relation between the electrostatic potential and the charge density at any point is given by Poisson's equation, based on Coulomb's law, and if this may be assumed to apply to ions, then

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial^2 \psi}{\partial z^2} = -\frac{4\pi \rho_e}{D}, \quad (117)$$

where x , y and z are the rectangular coordinates of the given point and D is the dielectric constant of the medium. Converting to polar coordinates, and making use of the fact that the terms containing $\partial \psi / \partial \theta$ and $\partial \psi / \partial \phi$ will be zero, since the distribution of potential about any point in the electrolyte is spherically symmetrical, (117) becomes

$$\frac{1}{r^2} \cdot \frac{\partial}{\partial r} \left(r^2 \frac{\partial \psi}{\partial r} \right) = -\frac{4\pi \rho_e}{D},$$

and introducing (116),

$$\frac{1}{r^2} \cdot \frac{\partial}{\partial r} \left(r^2 \frac{\partial \psi}{\partial r} \right) = \frac{4\pi\epsilon^2}{DkT} \psi \Sigma n_i z_i^2 = \kappa \psi, \quad (118)$$

where

$$\kappa = \sqrt{\frac{4\pi\epsilon^2 \Sigma n_i z_i^2}{DkT}}. \quad (119)$$

The general solution of the differential equation (118), whereby ψ may be expressed in terms of the distance r from the given ion, is

$$\psi = \frac{Ae^{-\kappa r}}{r} + \frac{A'e^{\kappa r}}{r}, \quad (120)$$

where A and A' are integration constants. Since ψ becomes zero as r increases to infinity, A' must be zero; further, A must be equal to $\pm z_i\epsilon/D$, so as to satisfy the condition that in a very dilute solution, when κ is very small, the potential near an ion is due to the ion alone considered as a point charge, i.e., $\pm z_i\epsilon/Dr$ according as the ion is positively or negatively charged.* Hence (120) becomes

$$\psi = \frac{\pm z_i\epsilon}{D} \cdot \frac{e^{-\kappa r}}{r} = \frac{\pm z_i\epsilon}{Dr} + \frac{\mp z_i\epsilon}{Dr} (1 - e^{-\kappa r}). \quad (121)$$

If κr is small, i.e., in dilute solution, $1 - e^{-\kappa r}$ is almost equal to κr and so

$$\psi = \frac{\pm z_i\epsilon}{Dr} + \frac{\mp z_i\epsilon\kappa}{D}. \quad (122)$$

In the absence of other ions the potential at a distance r from a given ion is $\pm z_i\epsilon/Dr$, as noted above, and hence the second term on the right-hand side must be that due to the oppositely charged atmosphere. Since this expression is independent of r , it may be assumed to hold when r is zero, so that the potential ψ_i on the ion itself due to its surrounding atmosphere is

$$\psi_i = \mp z_i\epsilon\kappa/D. \quad (123)$$

By means of (123) it is possible to ascribe a physical significance to κ . If the whole of the charge of the ionic atmosphere, i.e., $-z_i\epsilon$ for a positive ion, since it is equal and opposite to that of the ion itself, were placed at a distance $1/\kappa$ from the given ion the potential produced at it would be equal to $-z_i\epsilon\kappa/D$. It follows, therefore, that $1/\kappa$ may be regarded as the equivalent radius of the ionic atmosphere; it has the dimensions of a distance and is of the order of 10^{-8} cm. for ordinary solutions. As may be seen from (119), the actual value of $1/\kappa$ depends on the concentration of the solution and on the valences of the ions present.

The energy of a charged body, that is, the work expended in the process of charging, is half the product of its charge and potential; hence for an ion of charge $+z_i\epsilon$ the energy it possesses in virtue of its ionic atmosphere is, by (123), equal to $-\frac{1}{2}z_i^2\epsilon^2\kappa/D$. The corresponding energy for a 1 g. ion is obtained on multiplying by the Avogadro number N , so that

$$E_i = -Nz_i^2\epsilon^2\kappa/2D. \quad (124)$$

* The conclusion that A is equal to $z_i\epsilon/D$ is based on the tacit assumption that A is not a function of κ ; this is only the case for very dilute solutions.

The chemical potential of a particular ion in an ideal solution is (cf. p. 684)

$$\mu_i = \mu^0 + RT \ln x_i, \quad (125)$$

where x_i is its mole fraction in the given solution; for a nonideal solution, however,

$$\mu_i = \mu^0 + RT \ln a_i = \mu^0 + RT \ln x_i + RT \ln f_i, \quad (126)$$

where a_i is the activity and f_i the activity coefficient of the ionic species. The difference between (126) and (125), i.e., $RT \ln f_i$, is the difference in the free energy change accompanying the addition, or removal, of 1 g. ion of the given species from a large volume of real and ideal solution, respectively. This difference of free energy may be regarded as equivalent to the electrical energy of the ion due to its atmosphere; that is to say, the departure from ideal behavior is ascribed entirely to interionic forces. Hence, by (124),

$$RT \ln f_i = -Nz_i^2 e^2 \kappa / 2D; \quad (127)$$

$$\therefore \ln f_i = -Nz_i^2 e^2 \kappa / 2DRT. \quad (128)^*$$

Before substituting the value of κ from (119), n_i , the number of ions per cc., may be replaced by $Nc_i/1000$, where c_i is the number of g. ions per liter, and R/N may be substituted for k ; the natural logarithm in (128) may be converted to the common logarithm, so that

$$-\log f_i = \left(\frac{N^2 e^3 \sqrt{\pi/1000}}{2.3026R^{\frac{3}{2}}} \right) z_i^2 \frac{\sqrt{\sum c_i z_i^2}}{(DT)^{\frac{1}{2}}}. \quad (129)$$

The term in the large parentheses consists of universal constants only, and so may be replaced by A' ; further, comparison of $\sum c_i z_i^2$ with (110) shows that it is equal to twice the ionic strength, i.e., 2μ , and so (129) may be written

$$-\log f_i = \frac{A''}{(DT)^{\frac{1}{2}}} z_i^2 \sqrt{\mu}, \quad (130)$$

where A'' is $A'\sqrt{2}$. For a given solvent at a definite temperature D and T are constant and may be included with A'' to give a constant A , so that

$$-\log f_i = Az_i^2 \sqrt{\mu}. \quad (131)$$

The dielectric constant of water is 78.54 at 25° c., and inserting the known values of N , e , R and π , it is found that A is 0.509; the figure most frequently found in the literature of electrochemistry, however, is 0.505, based on the older value for the electronic charge. For most purposes it is probably sufficiently accurate to take A as 0.51 in water at 25° c.

Mean Activity of an Electrolyte.—Since the individual activity, or activity coefficient, of an ion cannot be determined experimentally, it is not possible to make a direct test of (130), but it is possible to relate the individual ion activities to the measurable mean activity of the ions. If one molecule of a binary electrolyte dissociates into ν ions of which ν_+ are cations and ν_- anions, then the mean activity coefficient f_{\pm} is related to

* The same result may be obtained by a more rigorous, but lengthier, argument.

the individual ionic activity coefficients f_+ and f_- by

$$f_{\pm} = \sqrt[{\nu_+ \nu_-}]{f_+^{\nu_+} f_-^{\nu_-}} \quad (132)$$

or

$$\log f_{\pm} = \frac{\nu_+ \log f_+ + \nu_- \log f_-}{\nu_+ + \nu_-}. \quad (133)$$

If the valences of the ions are z_+ and z_- respectively, then by the relationship between the valence and the number of ions, i.e., $\nu_+ z_- = \nu_- z_+$, it follows that

$$\log f_{\pm} = \frac{z_- \log f_+ + z_+ \log f_-}{z_+ + z_-}. \quad (134)$$

Inserting the value of $\log f_{\pm}$ for positive and negative ions, from (131) the result is

$$-\log f_{\pm} = A z_+ z_- \sqrt{\mu} \quad (135)$$

for a given solvent at a particular temperature. Equations (131) and (135) represent what is frequently called the **Debye-Hückel limiting law**, applicable to dilute solutions. According to this law the departure from ideal behavior in a given solvent is governed by the ionic strength of the medium and the valences of the ions of the electrolyte, but is independent of their chemical nature. The valence type of the electrolyte should thus be the essential factor in determining its activity coefficient at a given ionic strength.

Tests of the Debye-Hückel Theory.—Apart from the more exacting tests to be described shortly, it will be noted that (135) is in agreement with the empirical rule of Lewis and Randall (p. 956) that the mean activity coefficient in dilute solution is the same in solutions of the same ionic strength. Further, a number of workers had found that activity coefficients of uni-univalent electrolytes, determined by various methods, were a function of the square root of the concentration. For such substances the ionic strength is equal to the concentration, and so the results are in harmony with (135). The more careful studies made since 1923 have aimed particularly at verifying the numerical value of the constant A , since this provides a more rigid test of the theory than mere general agreement. Some of the methods used for the determination of activity coefficients, and the results obtained are described below.

Relationship between Activity Coefficients.—The derivation of (131) is based on the assumption that an ideal solution obeys Raoult's law or Henry's law [cf. equation (125)]; hence, the activity coefficient derived from the Debye-Hückel treatment is f_x (cf. p. 685), sometimes called the **rational activity coefficient**. In connection with solutions of electrolytes it is more convenient to use the activity coefficients f_c and f_m , referred to the ideal *dilute* solution; the latter, as already seen, is usually represented by the symbol γ . For these coefficients, the concentrations are given in

terms of g. ions or moles per liter, or molalities, i.e., moles per 1000 g. of solvent, respectively.

The concentration of a particular solution may be stated as x mole fraction of solute, c moles per liter, or molality m , and it can be readily shown that

$$x = \frac{0.001cM_1}{\rho - 0.001cM_2 + 0.001cM_1} = \frac{0.001mM_1}{1 + 0.001mM_1},$$

where ρ is the density of the solution and M_1 and M_2 are the molecular weights of solvent and solute, respectively. At very low concentrations the three quantities x_0 , c_0 and m_0 are thus related by

$$x_0 = 0.001c_0M_1/\rho_0 = 0.001m_0M_1,$$

where ρ_0 is the density of the solvent. If a is the activity at an appreciable concentration and a_0 that at infinite dilution, the ratio a/a_0 must be the same irrespective of the chosen standard state; since in each case it is postulated that the activity coefficient is unity at infinite dilution (p. 685), it follows that

$$\frac{a}{a_0} = \frac{f_x x}{x_0} = \frac{f_c c}{c_0} = \frac{\gamma m}{m_0},$$

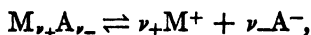
and these together with the expressions connecting x , c and m give

$$f_x = f_c(\rho - 0.001cM_2 + 0.001cM_1)/\rho_0 = \gamma(1 + 0.001mM_1).$$

These relationships between the three activity coefficients hold for any single species, but for the *mean* value for an electrolyte it is necessary to introduce ν , the number of ions produced by one molecule of electrolyte when it dissociates, in the $0.001cM_1$ and $0.001mM_1$ terms. It will be immediately evident that f_x and f_c must be almost identical, and that γ cannot differ appreciably from the other coefficients, for solutions more dilute than about 0.1 N.

Since most solutions used in electrochemical work are reasonably dilute the three activity coefficients may be assumed, for most purposes, to be identical. It is the general custom, however, to employ f_i , without the suffix, in connection with equilibria in electrolytes when the activity of a particular ionic species is required; thus $a_i = c_i f_i$, where c_i is the *actual* ionic concentration, due allowance being made for incomplete dissociation, if necessary. On the other hand, γ is most often used in connection with concentration cells when the activity of an ion is expressed as $m\gamma_i$, where m is the *total* molality of the electrolyte with no correction for incomplete dissociation; for this reason γ is sometimes called the *stoichiometric activity coefficient*.

Activities from Solubility Measurements.—In a saturated solution there is an equilibrium between the solid solute and the substance in solution; since the chemical potential of the former is constant, so also will be that of the latter at a given temperature. If one molecule of the substance dissociates into ν_+ positive and ν_- negative ions, thus



then by (16), p. 820, the chemical potential of the undissociated molecules must be equal to the sum of the potentials of the ions; either of these quantities may be taken as the chemical potential of the electrolyte in the solution. In the case under consideration, therefore, the chemical potential μ may be represented by

$$\mu = \nu_+ \mu_{M^+} + \nu_- \mu_{A^-} = \nu_+ (\mu_{M^+}^0 + RT \ln a_{M^+}) + \nu_- (\mu_{A^-}^0 + RT \ln a_{A^-}). \quad (136)$$

Since $\mu_{M^+}^0$ and $\mu_{A^-}^0$ are constant at a given temperature, and μ is constant as long as the solution is saturated, it follows that

$$\nu_+ \ln a_{M^+} + \nu_- \ln a_{A^-} = \text{constant} \quad (137)$$

or

$$(a_{M^+})^{\nu_+} (a_{A^-})^{\nu_-} = \text{constant } (K_s). \quad (138)$$

The constant K_s is the **activity solubility product**, and equation (138) expresses the solubility product principle, derived in a less exact manner by W. Nernst (1889). The activities may be written as the product of the respective concentrations and activity coefficients, so that

$$(c_+ f_+)^{\nu_+} (c_- f_-)^{\nu_-} = K_s \quad (139)$$

for a saturated solution. If S is the solubility of the salt in moles per liter, c_+ is equal to $\nu_+ S$ and c_- to $\nu_- S$, and hence (139) becomes

$$K_s = (\nu_+^{\nu_+} \nu_-^{\nu_-}) S^{\nu} f_{\pm}^{\nu}, \quad (140)$$

where ν is equal to the total number of ions, i.e., $\nu_+ + \nu_-$, produced from one molecule upon ionization and f_{\pm} is the mean activity coefficient [cf. equation (132)]. It follows therefore, from (140), since ν_+ and ν_- are constants for the given electrolyte, that

$$S f_{\pm} = \text{constant}. \quad (141)$$

If two solutions, which may contain added salts, are designated I and II, it follows from (141) that

$$S_I (f_{\pm})_I = S_{II} (f_{\pm})_{II} = \text{constant}, \quad (142)$$

a result which may be employed to determine the mean activity coefficient of a sparingly soluble salt. This is done by making solubility measurements in the presence of added salts at various ionic strengths, and extrapolating to infinite dilution. Since f_{\pm} is then unity, the extrapolated solubility gives the constant of (142); once this is known the mean activity coefficient can be evaluated from solubility measurements in any solution.

For the purpose of testing the Debye-Hückel theory (142) may be written

$$S/S_0 = f_{0\pm}/f_{\pm} \quad \text{or} \quad \log S/S_0 = \log f_{0\pm} - \log f_{\pm}, \quad (143)$$

where S_0 and S refer to the saturation solubilities in pure water and in the presence of various added salts. Introducing the mean activity

coefficients given by (135) it is seen that

$$\log S/S_0 = Az_+z_-(\sqrt{\mu} - \sqrt{\mu_0}). \quad (144)$$

Since μ_0 , which is determined by the solubility in pure water, is constant it follows from (144) that the plot of $\log S/S_0$ for a given saturating salt against $\sqrt{\mu} - \sqrt{\mu_0}$, should approach a straight line in dilute solutions; its slope should be Az_+z_- , irrespective of the nature of the neutral electrolytes present. The interesting results on the solubilities of the sparingly soluble cobaltammine compounds, viz., $[\text{Co}(\text{NH}_3)_4(\text{NO}_2)_2(\text{CNS})][\text{Co}(\text{NH}_3)_2(\text{NO}_2)_2(\text{C}_2\text{O}_4)]$, which is a 1-1 valence type (I); $[\text{Co}(\text{NH}_3)_4(\text{C}_2\text{O}_4)]_2\text{S}_2\text{O}_8$, belonging to the 1-2 type (II); and $[\text{Co}(\text{NH}_3)_6][\text{Co}(\text{NH}_3)_2(\text{NO}_2)_2(\text{C}_2\text{O}_4)]_3$, of the 3-1 type (III), obtained by J. N. Brønsted and V. K. LaMer (1924) are plotted in Fig. 239. The measurements

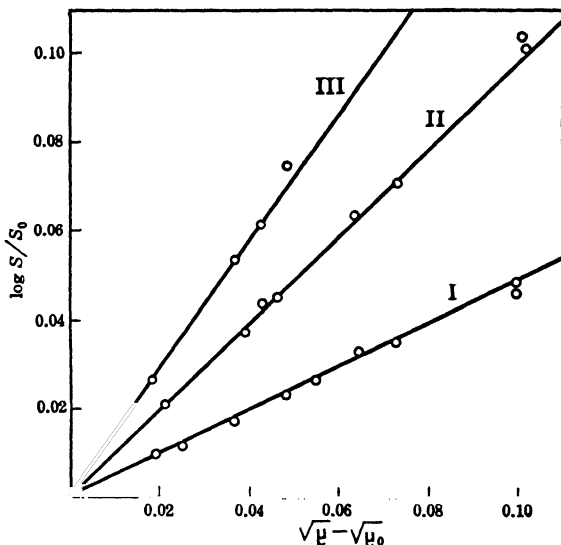


FIG. 239. Variation of solubility with ionic strength

were made in the presence of a variety of added salts, and the agreement with the Debye-Hückel theory in dilute solutions is very striking. If both ions of the saturating salt have high valences discrepancies are noted, especially if the added electrolytes also contain polyvalent ions. Solubility measurements have also been made in various nonaqueous media, so as to test the dependence of the constant A on the dielectric constant; the results are, on the whole, in general harmony with the theoretical requirements in very dilute solutions.

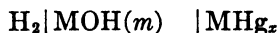
Activities from E.M.F. Measurements.—It was seen on p. 942 that the E.M.F. of the cell



could be expressed, by (91), in terms of a constant E^0 , the molality of the solution of hydrochloric acid and the mean activity coefficient of the electrolyte. It was also shown that E^0 could be evaluated by suitable extrapolation. Since E^0 is known it is clearly possible to calculate γ_{\pm} at any molality from the measured E.M.F. of the cell containing hydrochloric acid at that molality. The method can be extended to any electrolyte provided it is possible to devise a cell in which each electrode is reversible with respect to one of the ions. For an alkali chloride, for example, an amalgam cell



where MHg_x is the alkali-metal amalgam, may be used, and for alkaline hydroxides the cell



may be employed, since the hydrogen electrode can be regarded as being reversible with respect to OH^- ions (see p. 991).

The method described above requires that accurate measurements in dilute solution should be available; otherwise the extrapolation to infinite dilution, for the evaluation of E^0 , is unreliable. If this is not possible, then use may be made of the E.M.F. of the concentration cell without transference obtained by combining two of the cells described above. For a uni-univalent electrolyte

$$E = \frac{2RT}{F} \ln \frac{m_1 \gamma_1}{m_2 \gamma_2}, \quad (145)$$

as shown on p. 927, where γ_1 and γ_2 are the mean ionic activities in the solutions of molality m_1 and m_2 . If γ_1 is known from other measurements, e.g., from freezing point determinations or by comparison with another electrolyte of the same type, γ_2 can be evaluated from the E.M.F. of the cell.

A procedure has also been described for the evaluation of activity coefficients from measurements on concentration cells with transference; this requires, however, an accurate knowledge of transference numbers over the concentration range employed in the cells. Where the necessary data are available reliable activity coefficients have been obtained in this manner (D. A. MacInnes, *et al.*, 1935 *et seq.*).

A selection of activity coefficients derived from E.M.F. measurements at 25° c is given in Table 161. For molalities at least up to 0.05 the mean activity coefficients are almost independent of the nature of the electrolyte, provided it is of the 1-1 valence type; this is in accordance

TABLE 161. ACTIVITY COEFFICIENTS FROM E.M.F. MEASUREMENTS

<i>m</i>	LiCl	NaCl	KCl	KBr	KOH
0.001	—	0.966	0.965	—	0.989
0.005	—	0.928	0.926	—	0.954
0.01	0.901	0.903	0.899	—	0.920
0.05	0.819	0.821	0.815	0.816	0.822
0.1	0.779	0.778	0.764	0.765	0.789
0.5	0.725	0.678	0.644	0.653	0.750
1.0	0.757	0.658	0.597	0.612	0.760
2.0	0.919	0.670	0.569	0.589	—
3.0	1.174	0.714	0.571	0.590	1.062

with the requirements of the Debye-Hückel theory. The limiting slope of the plot of $-\log f_{\pm}$ against \sqrt{u} is very close to that required by (135).

Activities from Freezing Point Measurements.—According to the deduction on p. 690 the activity a_2 of a solute in a dilute solution is related to the depression of the freezing point ΔT_f by the expression

$$d \ln a_2 = d(\Delta T_f)/mK_f, \quad (146)$$

where K_f is the molal depression constant. As seen above [equation (136)] the chemical potential of a dissolved salt may be taken as the sum of the contributions of the two ions; thus for a uni-univalent electrolyte,

$$\mu_2 = \mu_+ + \mu_- = (\mu_+^0 + RT \ln a_+) + (\mu_-^0 + RT \ln a_-), \quad (147)$$

where μ_2 is the chemical potential of the solute, which may also be expressed in the form

$$\mu_2 = \mu_2^0 + RT \ln a_2. \quad (148)$$

The standard states of the ions are defined so as to make the activities equal to the molalities at infinite dilution (cf. p. 685) but that of the solute as a whole is chosen so that

$$\mu_2^0 = \mu_+^0 + \mu_-^0. \quad (149)$$

Combination of (147), (148) and (149) gives $a_2 = a_+a_-$, and since the mean activity of the ions a_{\pm} is equal to $\sqrt{a_+a_-}$, it follows that $a_{\pm} = \sqrt{a_2}$. In general for a solute dissociating into ν ions,

$$a_{\pm} = \sqrt[\nu]{a_+^{\nu_+}a_-^{\nu_-}} = \sqrt[\nu]{a_2}; \quad (150)$$

$$\therefore \ln a_2 = \nu \ln a_{\pm} \quad (150a)$$

and by differentiation

$$d \ln a_2 = \nu d \ln a_{\pm}. \quad (151)$$

Substitution in (146) gives

$$d \ln a_{\pm} = d(\Delta T_f)/\nu m K_f. \quad (152)$$

A function j may be defined by the expression

$$j = 1 - \Delta T_f/\nu m K_f, \quad (153)$$

so that j becomes zero at infinite dilution, when ΔT_f is equal to $\nu m K_f$. Differentiating (153) and simplifying, it is found that

$$\frac{d(\Delta T_f)}{\nu m K_f} = (1 - j) \frac{dm}{m} - dj \quad (154)$$

and hence from (152),

$$d \ln a_{\pm} = (1 - j) \frac{dm}{m} - dj. \quad (155)$$

Subtract $d \ln m$ from both sides of this equation, and there is obtained

$$d \ln \frac{a_{\pm}}{m} = d \ln \gamma_{\pm} = -j \frac{dm}{m} - dj \quad (156)^*$$

$$\therefore \ln \gamma_{\pm} = - \int_0^m \frac{j}{m} dm - j, \quad (157)$$

remembering that j is 0 when m is 0. If the lowering of the freezing point (ΔT_f) at any molality (m) is known, then substitution in (153) gives the corresponding value of j . If $-j/m$ is plotted against m , the area under the curve from 0 to m gives the integral in (157) and so $\ln \gamma_{\pm}$ can be obtained. The graphical method is somewhat tedious, and G. N. Lewis (1919) expressed j in terms of m by the empirical equation $j = \beta m^{\alpha}$, the constants α and β being determined from the experimental data; substitution for j in (157) and integration, gives

$$\ln \gamma_{\pm} = - \frac{\beta}{\alpha} m^{\alpha} - \beta m^{\alpha}. \quad (158)$$

Since α and β are known, $\ln \gamma_{\pm}$ can be evaluated at any desired molality, provided the solution is dilute. Since the activity coefficients are determined at the freezing point of the solution, corrections must also be included if the values at 25° c. are required (see p. 687). Some of the results obtained from freezing point measurements are recorded in Table 162; these have been specially chosen to represent different valence types.

TABLE 162. ACTIVITY COEFFICIENTS AT 25° C. FROM FREEZING POINT DATA

m	KCl	K ₂ SO ₄	BaCl ₂	Pb(NO ₃) ₂	CuSO ₄
0.0001	0.982	0.935	0.94	0.965	0.85
0.0005	0.970	0.885	0.89	0.922	0.75
0.005	0.940	0.75	0.77	0.773	0.50
0.01	0.903	0.69	0.72	0.694	0.40
0.05	0.821	0.505	0.568	0.465	0.216
0.1	0.779	0.421	0.501	0.369	0.158

Freezing point measurements have been made in a wide variety of solvents, e.g., liquid ammonia, acetic acid and cyclohexanol, in addition to water; the activity coefficients derived from them approach the behavior required by the limiting Debye-Hückel equation as the solutions become more dilute.²⁴

The Complete Debye-Hückel Equation.—In the simple Debye-Hückel treatment the approximation was made of regarding the ion as a point charge. This causes no considerable error in dilute solutions, but at concentrations of the order of 0.01 molar the effect becomes noticeable. For solutions of appreciable concentration an allowance for the size of the ions is necessary, and this may be made by introducing the factor $1/(1 + \kappa a)$ into (131) and (135), where α

* Comparison of this result with (93), p. 686, shows that the practical osmotic coefficient ϕ is equal to $1 - j$, i.e., to $\Delta T_f / \nu m K_f$.

is the average distance of approach of two oppositely charged ions; it is generally referred to as the average effective diameter of the ions of the electrolyte. Making this correction, it is found that

$$-\log f_{\pm} = \frac{Az_+z_- \sqrt{\mu}}{1 + \kappa a} \quad (159)$$

Since κ involves the square root of the ionic strength, this can be written

$$-\log f_{\pm} = \frac{Az_+z_- \sqrt{\mu}}{1 + aB \sqrt{\mu}}, \quad (160)$$

where B is made up of universal constants, the dielectric constant and the temperature; for water at 25° c. it is 3.29×10^7 . Equation (160) is an improvement on the limiting form, (135), but in relatively concentrated solutions it is evident that another term must be introduced. It is seen from Table 161 that although activity coefficients decrease at first with increasing concentration of electrolyte, as required by the Debye-Hückel equations, a point is always reached at which the activity coefficient commences to increase. It is evident, therefore, that the actual behavior of solutions can only be represented by subtracting * from (160) a term which increases with concentration. E. Hückel (1925) considered that owing to the attraction between the ions and the dipolar molecules of solvent, there is a tendency for the latter to displace other ions to some extent from the vicinity of any particular ion. This is equivalent to a salting-out effect (p. 700), and correction for it may be made by assuming a change in the dielectric constant by an amount proportional to the ionic concentration. The final form of the Debye-Hückel equation is

$$-\log f_{\pm} = \frac{Az_+z_- \sqrt{\mu}}{1 + aB \sqrt{\mu}} - C\mu. \quad (161)$$

Unfortunately an exact test of this equation is not possible, since both a and C must be derived empirically, but it is certainly true that by utilizing reasonable values of the average diameter a , e.g., about 2 to 5×10^{-8} cm., the behavior of electrolytes can be represented in a satisfactory manner. The general effect of the corrections may be seen from Fig. 240; the straight line I represents the limiting Debye-Hückel law, curve II shows the effect of introducing the correction for ionic size, with $a = 2.5 \times 10^{-8}$ cm., and finally III indicates the influence of the salting-out factor. Incidentally curve III duplicates quite closely the variation of the activity coefficient of sodium chloride up to quite high concentrations. For many purposes (161) is

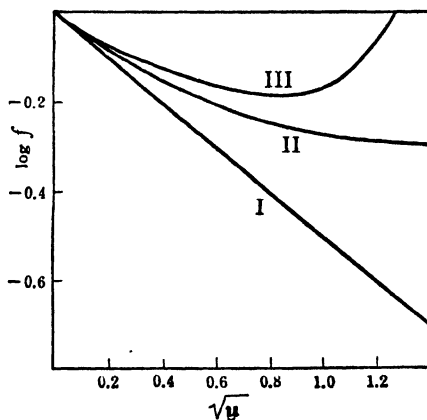


FIG. 240. Simple (I) and extended (II and III) Debye-Hückel equations

* Subtraction is necessary since (160) gives $-\log f_{\pm}$.

written in the more convenient form

$$-\log f_{\pm} = Az_+z_- \sqrt{\mu} - C'\mu, \quad (162)$$

where C' is approximately equal to $z_+z_-aAB + C$.

Thermal Properties of Electrolytes.—The electrical work required to separate the ions of 1 mole of electrolyte to an infinite distance from each other, which is equivalent to the free energy of diluting the solution to infinity, can be shown to be two-thirds the electrical energy of the ions as given by (124). In order to obtain the integral heat of dilution, that is, the heat evolved when the solution containing 1 mole is mixed isothermally with an infinite bulk of solvent, it is necessary to apply the Gibbs-Helmholtz equation; the result for a dilute solution of a uni-univalent electrolyte, at constant pressure, is

$$\Delta H_{\infty 0} = ART \left[1 + \frac{T}{D} \left(\frac{\partial D}{\partial T} \right)_P + \frac{T}{3V} \left(\frac{\partial V}{\partial T} \right)_P \right] \sqrt{\mu}, \quad (163)$$

where A has the same significance as in the Debye-Hückel equation, and V is the volume of solvent containing 1 mole of electrolyte; $\partial D/\partial T$ is the temperature coefficient of the dielectric constant and $(\partial V/\partial T)/V$ is the coefficient of thermal expansion of the solvent. Inserting the known values, and also those for the constants A and R , it follows that at 25° c. for an aqueous solution,

$$\Delta H_{\infty 0} = -495 \sqrt{c} \text{ cal. per mole}, \quad (164)$$

the negative sign arising because $\partial D/\partial T$ is negative. This result implies that heat is evolved in the dilution of a dilute solution, and that the amount should be a linear function of the square root of the concentration. Accurate experimental results are difficult to obtain, but on the whole they may be said to support the interionic attraction theory. In relatively concentrated solutions marked deviations occur, and the heat of dilution changes sign.

Extension of the Debye-Hückel Treatment.—In the calculation of the electric density in the vicinity of an ion it was assumed that $z_i e \psi / kT$ was small (p. 957), so that all terms after the first in the exponential series could be neglected. This approximation is not justifiable for ions of high valence or small radius, and so T. H. Gronwall and V. K. La Mer (1928, 1931) used the complete series and then carried out an approximate integration of the Poisson equation. The result was similar to (159) with the addition of a series of terms, all but the first two being negligibly small, which are functions of κa . If a/z^2 for a symmetrical electrolyte is greater than about 1.5×10^{-8} cm. the additional terms are negligible in aqueous solution. Hence, for a uni-univalent electrolyte the Debye-Hückel treatment will apply if the mean ionic diameter is greater than about 1.5 Å., but for a bi-bivalent salt the diameter must exceed 6 Å. Since the diameters of most ions are greater than 1.5 Å. but less than 6 Å. the simple equation will hold for ions of low valence, but not for high valence electrolytes. For solvents of low dielectric constant the Gronwall-La Mer terms become negligible only if the diameters are larger than in water, and so the deviations from the Debye-Hückel equation are more marked. The extended equation, involving the extra terms, has been tested quantitatively, and there is little doubt that for solutions of appreciable concentration it is an improvement on the original

treatment. For very dilute solutions the correction terms become negligible, since κ is then small, and (131) and (135) are applicable.*

An alternative extension of the Debye-Hückel theory has been considered by N. Bjerrum (1926). He suggested that if the sum of the radii of the two ions of an electrolyte exceeds the minimum value given by

$$r_{\min.} = z_+ z_- \epsilon^2 / 2DkT, \quad (165)$$

then the theory may be regarded as applicable, but when it is less than this the ions should be considered as being associated in pairs, in which state they contribute nothing to the electrical energy of the ion due to its atmosphere. The general conclusion is in agreement with Gronwall and La Mer, that the smaller the ions, the higher their valence and the lower the dielectric constant, the more likely are the results to show deviations from the approximate treatment. According to (165) $r_{\min.}$ is 3.5 Å. for a uni-univalent salt in water at 25° C., and if the sum of the ionic radii is smaller than this there will be a tendency for ion-pairs to form; the electrolyte will then behave as if it were incompletely dissociated. If the sum of the ionic radii is 6.8 Å., then it follows from (165) that dissociation of a completely ionized electrolyte should be complete in solvents of dielectric constant exceeding 42, at ordinary temperatures. It has been calculated from conductance measurements on tetraisoamylammonium nitrate in dioxane-water mixtures (cf. p. 897), for which the data are in accord with a mean diameter of 6.8 Å., that there is no appreciable ion-pair formation when the dielectric constant exceeds 41 (R. M. Fuoss and C. A. Kraus, 1933).

An important extension of the concept of ion-association has been made by R. M. Fuoss and C. A. Kraus (1933 *et seq.*); they considered the possibility of the formation of triple ions in addition to ion-pairs, and as a result were able to derive theoretically the empirical rule of Walden (p. 896) concerning the minimum of equivalent conductance. It can be seen, in a qualitative manner, that as the concentration of an electrolyte increases the extent of ion-association to form pairs, viz., $+-$, will increase, so that the equivalent conductance diminishes. With increasing concentration, however, triple ions, viz., $+--+$ and $-+-$, will commence to form, but as these have a resultant charge they will contribute toward the conductivity which consequently commences to rise again. The smaller the dielectric constant of the medium the greater the tendency for the ions to associate in pairs or in triplets and hence the lower the concentration at which the conductance minimum occurs.²⁵

Weak Electrolytes and the Debye-Hückel Theory.—Although the interionic action theory was developed primarily in an attempt to solve the problem of strong electrolytes, which were assumed to be completely ionized, the conclusions are equally applicable to weak and intermediate electrolytes. The only adjustment which it is necessary to make in the Debye-Hückel equations is that in evaluating the ionic strength μ the actual ionic concentrations must be employed. For a strong electrolyte, in which ionization is complete, the term c ; in the expression $\frac{1}{2} \sum c_i z_i^2$ may be derived directly from the concentration of the solute, but for a weak

* Some doubt has been expressed concerning the complete consistency of the Gronwall-La Mer treatment.

or intermediate electrolyte it is necessary in addition to introduce the degree of dissociation. For a uni-univalent compound, for example, c_i for each ion is equal to αc , where c is the electrolyte concentration; hence the ionic strength is αc , instead of c for a strong electrolyte.

If the ionic strength is not too high the activity coefficient of the non-ionized molecules MA , that is, f_{MA} , may be taken as unity; combination of (108) with (109) then leads to the relationship $K = k(f_{M^+} \times f_{A^-})$ between the "thermodynamic" dissociation constant K , and the "classical" quantity k . If the simple Debye-Hückel equation (131) for the activity coefficient of each ion is introduced in the form $\log f_i = -A \sqrt{\alpha c}$, since z_i is unity for both ions and ν is equal to αc , it follows that

$$\log k = \log K + 2A \sqrt{\alpha c}. \quad (166)$$

The value of α to be used in this equation is not the conductance ratio Λ/Λ_0 , but the quantity previously represented by Λ/Λ' , where Λ' is the equivalent conductance corrected for interionic forces either empirically (p. 902) or by the use of the Onsager equation (p. 907). It is clear, therefore, from (166) that the classical dissociation constant, as defined by (109), i.e., $k = \alpha^2 c / (1 - \alpha)$, even when α has been appropriately corrected, should increase with increasing concentration; this is in agreement with experiment, as shown by the results in Table 160. By taking A as 0.51 at 25° K. it is a simple matter to apply the activity correction $2A \sqrt{\alpha c}$, and when this is done the first six solutions in the table give dissociation constants K which are almost exactly 1.750×10^{-5} . In the more concentrated solutions the values fall off somewhat, because of the incomplete nature of the simple Debye-Hückel equation, but the necessary correction could be made empirically if the activity coefficients are expressed by (162), so that (166) becomes

$$\log k = \log K + 2A \sqrt{\alpha c} - \alpha c C'. \quad (166a)$$

The reliability of (166) can be tested in another manner: by inserting the equation for k given above, it is seen that

$$\log \alpha^2 c / (1 - \alpha) = \log K + 2A \sqrt{\alpha c}, \quad (167)$$

so that the plot of $\log \alpha^2 c / (1 - \alpha)$ against $\sqrt{\alpha c}$ for solutions of different concentrations should be a straight line, its slope being $2A$. In one or other of these ways the value of the Debye-Hückel activity correction has been confirmed for weak and moderately strong acids, as well as for bi-bivalent salts, e.g., sulfates and oxalates, which appear to be undissociated to an appreciable extent.

Solubility Product and Solubility.—In its original form, as given by Nernst, the solubility product principle stated that for a saturated solution of a sparingly soluble electrolyte $M_{r+}A_{r-}$,

$$(c_{M^+})^{r+}(c_{A^-})^{r-} = \text{constant } (k_s), \quad (168)$$

the constant being called the solubility product. According to the activity principle, however, the product $(a_{M^+})^{+}(a_{A^{-}})^{-}$ should be constant [equation (138)], and hence (168) can be correct only when the activity coefficients are constant. In spite of this limitation the simple solubility product principle, with ionic concentrations, has proved of value in analytical chemistry, and in many instances gives results which are also quantitatively correct. It is seen from (168) that the addition of a salt with a common ion should decrease the concentration of the other ion of the sparingly soluble salt; there should, therefore, be a reduction in the solubility of the latter. The extent of the decrease may be calculated in the following manner. Let S_0 be the solubility of a uni-univalent salt, e.g., silver chloride, in water; then if it may be regarded as completely dissociated at the very low concentration of the saturated solution, $c_{Ag^+} = c_{Cl^-} = S_0$. According to the simple solubility product principle,

$$k_s = c_{Ag^+} \times c_{Cl^-} = S_0^2. \quad (169)$$

Suppose x g. equiv. of a soluble chloride, e.g. potassium chloride, assumed to be completely dissociated, is added to a liter of solution; this adds x g. ions of chloride ions to the solution, and the solubility of the silver chloride is decreased to S . The value of c_{Ag^+} is now S , whereas c_{Cl^-} is $S + x$; hence,

$$k_s = c_{Ag^+} \times c_{Cl^-} = S(S + x) = S_0^2. \quad (170)$$

Solving for S , it is found that

$$S = -\frac{1}{2}x + \sqrt{\frac{1}{4}x^2 + S_0^2}, \quad (171)$$

so that if S_0 is known the solubility in the presence of the added common ion can be calculated. The solubilities of various sparingly soluble salts derived in this manner are in fair agreement with the observed values, but where discrepancies are observed they can be attributed to the neglect of the activity coefficients. For example, it is almost invariably true that the addition of a common ion at first diminishes the solubility, but when relatively large amounts are present the solubility commences to increase. By the activity principle the product $a_{M^+}a_{A^{-}}$, for a uni-univalent salt, must remain constant; that is to say, $(c_{M^+}f_{M^+})(c_{A^{-}}f_{A^{-}})$ should be constant. If an electrolyte with a common ion, e.g., M^+ , is added, then $c_{A^{-}}$ will decrease to maintain the product constant. It must be remembered that at the same time the ionic strength increases and the activity coefficients decrease. A point may be reached, therefore, at which $c_{A^{-}}$ will have to *increase* as a result of an increase of c_{M^+} because of the large accompanying decrease in f_{M^+} and $f_{A^{-}}$. As is to be expected this effect is more easily observed with ions of high valence, e.g., lanthanum iodate in the presence of lanthanum nitrate.

The presence of a neutral salt without a common ion often results in an increase of solubility of a sparingly soluble electrolyte, and this may be followed by a decrease at high salt concentrations. Although the

simple solubility (concentration) product principle is unable to account for these results, they are readily explicable by the activity concept. The addition of the neutral salt increases the ionic strength of the medium and so $f_M + f_{A^-}$ decreases; hence $c_M + c_{A^-}$, and thus the solubility, must increase to maintain the solubility (activity) product. At high ionic strengths, however, the activity coefficients may begin to increase (p. 967); the solubility of the saturating salt will consequently decrease.

Determination of Solubility Products.—If a salt is very sparingly soluble, its saturated solution in water is so dilute that the salt may be assumed to be completely ionized and the activity coefficients of the ions equal to unity. Under these conditions

$$\begin{aligned} k_s &= (c_M^+)^{\nu_+} (c_{A^-})^{\nu_-} = (\nu_+ S)^{\nu_+} (\nu_- S)^{\nu_-} \\ &= (\nu_+)^{\nu_+} (\nu_-)^{\nu_-} S^{\nu} \end{aligned} \quad (172)$$

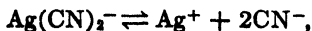
where S is the solubility of the salt in g. moles per liter, and ν is equal to the sum of ν_+ and ν_- ; hence if the solubility is known the solubility product may be evaluated. For this purpose the conductance method (p. 899) may be utilized.

Another principle used in the determination of solubility products is to measure the concentration of one of the ions by a reversible electrode; the electrode $M | MA$ (solid) NaA soln. is set up and its potential measured by combining it with a reference electrode. If the standard potential of the metal is known then the apparent * activity of the M^+ ions in the solution may be derived from the usual equation (p. 938); thus,

$$E = E^0 - RT/zF \ln a_{M^+}. \quad (173)$$

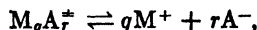
From the concentration of the NaA solution the approximate activity of the A^- ions can be obtained, by assuming it to be equal to the mean activity of the electrolyte. The product of the apparent activities of the ions M^+ and A^- gives the solubility product. Other E.M.F. methods have been described but they involve the same fundamental principle; in some the difficulty concerning the actual ionic activities is overcome, theoretically, by suitable cells.²⁶

Complex Ions.—The addition of a certain molecular or ionic species sometimes causes a large increase in the solubility of a sparingly soluble substance. For example, silver cyanide is readily soluble in solutions of CN^- ions; silver chloride and oxide dissolve in aqueous ammonia; and mercuric iodide is soluble in potassium iodide solution. In each case the increase of solubility is due to the removal of the simple ions in the form of complex ions, viz., $Ag(CN)_2^-$, $Ag(NH_3)_2^+$, and HgI_4^{--} , respectively; more of the insoluble substance must then dissolve to maintain the solubility product. It is probable that there is always an equilibrium between the complex ion and the simple ions from which it forms, e.g.,



* True single ion activities could only be obtained if liquid junction potentials were completely eliminated (see p. 945, footnote).

in spite of the great stability, in this case, of the complex ion. Migration experiments show that virtually the whole of the silver is present in the anion, but the precipitation of silver sulfide on addition of hydrogen sulfide suggests that the solution contains some, even if few, free silver ions. In general the equilibrium for the complex ion $M_qA_r^{\pm}$ may be written



and application of the law of equilibrium gives

$$K_i = \frac{a_{M^+}^q \times a_{A^-}^r}{a_{M_qA_r^{\pm}}}, \quad (174)$$

or using concentrations, as is generally done for simplicity,

$$k_i = c_{M^+}^q \times c_{A^-}^r / c_{M_qA_r^{\pm}}. \quad (175)$$

The constant, K_i or k_i , is the **instability constant** of the complex ion, for it gives a measure of its tendency to dissociate into simple ions. The reciprocal of k_i , called the "stability constant," was originally defined by G. Bodländer (1903).

Two methods have been chiefly used for determining the formulae and instability constants of complex ions: the first involves solubility measurements and may be described by reference to a special case, the silver-ammonia complex ion. Suppose m g. ions of silver combine with n moles of ammonia to form the cation $Ag_m(NH_3)_n^{m+}$; the approximate instability constant, using concentrations instead of activities, is then

$$k_i = c_{Ag^+}^m \times c_{NH_3}^n / c_{comp.}, \quad (176)$$

where $c_{comp.}$ represents the concentration of the complex ion. If a solution of ammonia is saturated with silver chloride, then by the approximate solubility product principle $c_{Ag^+} \times c_{Cl^-}$ is equal to k_s , and so $c_{Ag^+} = k_s / c_{Cl^-}$; hence

$$k_i = k_s^m \times c_{NH_3}^n / c_{Cl^-}^m \times c_{comp.} \quad (177)$$

Since the normal solubility of the silver chloride is small, the whole of the silver in solution may be assumed to be in the form of the complex ion; further, the chloride ion concentration will be the same as that of the latter. If c is the total silver concentration in the solution, then $c_{comp.} = c_{Cl^-} = c$, and (177) becomes

$$k_i = k_s^m \times c_{NH_3}^n / c^{m+1}; \quad (178)$$

$$\therefore c_{NH_3}^n / c^{m+1} = \text{constant}. \quad (179)$$

By means of (179) the ratio $n/(m+1)$ can be evaluated from a number of solubility measurements in ammonia solutions of different concentrations. If the solubility of the silver chloride is determined in the presence of an excess of chloride ions, then $c_{comp.}$ is not equal to c_{Cl^-} , and (177) is then

$$k_i = k_s^m \times c_{NH_3}^n / c_{Cl^-}^m \times c; \quad (180)$$

$$\therefore c_{NH_3}^n / c_{Cl^-}^m \times c = \text{constant}. \quad (181)$$

If in a series of experiments c_{NH_3} is kept constant, then $c_{Cl^-}^m \times c$ should be constant, and hence m can be determined.

The E.M.F. method described by G. Bodländer (1901) has been frequently used. Consider again the silver-ammonia ion; it follows from (176) that for two solutions, indicated by 1 and 2, at different concentrations

$$\frac{(c_{Ag^+})_1^m}{(c_{Ag^+})_2^m} = \frac{(c_{comp.})_1(c_{NH_3})_2^n}{(c_{comp.})_2(c_{NH_3})_1^n} \quad (182)$$

Suppose a concentration cell is made up with two silver electrodes in the two solutions, then the E.M.F. of the cell, excluding the liquid junction potential, is

$$E = \frac{RT}{F} \ln \frac{(c_{Ag^+})_1}{(c_{Ag^+})_2} = \frac{RT}{mF} \ln \frac{(c_{comp.})_1(c_{NH_3})_2^n}{(c_{comp.})_2(c_{NH_3})_1^n}, \quad (183)$$

using concentrations instead of activities. If the solutions both contain the same large excess of ammonia, then $(c_{NH_3})_1$ and $(c_{NH_3})_2$ can be regarded as equal; hence,

$$E = \frac{RT}{mF} \ln \frac{(c_{comp.})_1}{(c_{comp.})_2} \quad (184)$$

If the complex ion is stable then $c_{comp.}$ may be put equal to the total concentration c of silver in the solution, and so

$$E = \frac{RT}{mF} \ln \frac{c_1}{c_2} \quad (185)$$

It is thus possible to evaluate m from measurement of the E.M.F. of the cell, or by measuring each half-cell separately against a reference electrode. On the other hand, if two solutions with the same concentration of complex ion are taken, that is to say they contain the same total amount of silver, then (183) becomes

$$E = \frac{RT}{mF} \ln \frac{(c_{NH_3})_2^n}{(c_{NH_3})_1^n} \quad (186)$$

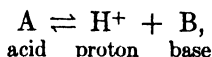
Since m is known the value of n can be obtained from the E.M.F. of the cell with different amounts of ammonia on each side. Although the methods have been described with reference to a particular complex ion they can be easily adapted to any other such ion.²⁷

ACIDS AND BASES

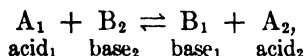
Acids and Bases.*—The definitions of an acid as a substance which yields hydrogen ions and a base as one giving hydroxyl ions, and of neutralization as the production of salt and water from an acid and a base, are reasonably adequate for aqueous solutions, but there are obvious limitations when nonaqueous media are concerned. As a result of various studies, particularly those on the catalytic influence of nonionized molecules and certain ions (Chapter XIII), the view was developed independently by J. N. Brønsted (1923) and T. M. Lowry (1923) that an

* G. N. Lewis (1938) has proposed a broad definition of a base as a substance capable of furnishing a pair of electrons to a bond, i.e., an electron donor, while an acid is able to accept a pair of electrons, i.e., an electron acceptor. Although this general definition is of great significance, it is more convenient from the electrochemical standpoint to use the somewhat more restricted definitions given here.

acid should be defined as a substance having a tendency to lose a proton,* and a base as one with a tendency to gain a proton. The relationship between an acid and a base may then be expressed in the form



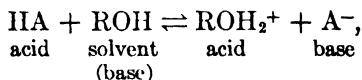
where the acid A and the base B, which differ by a proton, are said to be **conjugate** to one another. Every acid must, in fact, have its conjugate base, and every base its conjugate acid. The interaction of an acid A₁ of one system with a base B₂ of another, and in fact almost all types of acid-base reactions, may be represented by the equilibrium



where A₁ and B₁ are the conjugate acid and base of one system, and A₂ and B₂ are those of the other system. Actually A₁ possesses a proton more than B₁, while A₂ has a proton more than B₂; the reaction therefore involves the transfer of a proton from A₁ to B₂ in one direction, or from A₂ to B₁ in the other direction.

It is unlikely that free protons exist to any extent in solution, and so the acidic or basic functions of any species cannot become manifest unless the solvent molecules are themselves able to act as proton acceptors or donors, respectively; that is to say, the medium itself must have basic or acidic properties. The interaction between an acid or base and the basic or acidic solvent, respectively, is then similar to the equilibrium represented above. For convenience solvents are divided roughly into three categories according as the molecules are (a) proton acceptors, i.e., basic, or **protophilic**, (b) proton donors, i.e., acidic, or **protogenic**, or (c) neither donors nor acceptors, i.e., **aprotic**.

Acids.—Water and alcohols are good examples of protophilic solvents, since the oxygen atom of the molecule is readily able to accept a proton. The equilibrium of an acid in a liquid of this type, designated ROH, where R may also be H, is represented by

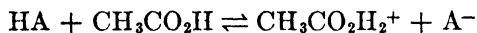


where ROH₂⁺ is the entity usually called the "hydrogen ion" in solution, although it is really a form of oxonium ion. Such ions are probably solvated further, but this fact may be ignored for the present. It has been seen on p. 898 that in aqueous solution the "hydrogen ion" is probably, at least, H₃O⁺, in agreement with the foregoing suggestion; this ion is sometimes called the hydroxonium or hydronium ion, but it is simpler to refer to it as the hydrogen ion. It must be remembered, never-

* The term "proton" as used here applies equally to the deuteron in deuterio-acids and deuterio-bases.

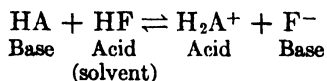
theless, that the so-called hydrogen ion is different in each solvent; as long as this is borne in mind, however, it is still convenient to use the symbol H^+ to represent it. Although the symbol is the same for different media the actual ions are not; it is a different oxonium ion, for example, in each hydroxylic solvent.

Since the hydrogen ion is an acid, in the general sense, and the anion A^- is a base, a certain position of equilibrium will always be attained in the interaction between an acid and a protophilic solvent, as shown above. The stronger the protophilic nature of the solvent the greater the extent to which the reaction goes from left to right. For this reason strong acids behave as almost equally strong in aqueous solution, because in each case the basic nature of water results in almost complete conversion into the H_3O^+ ion. In water, therefore, except in very concentrated solutions, any strong acid consists almost entirely of hydroxonium ions, and the strengths cannot be distinguished. In a weakly protophilic medium, however, such as when acetic acid is the solvent, the position is different. The equilibrium



cannot go very far to the right, and the extent of the formation of the $CH_3CO_2H_2^+$ (hydrogen) ions is dependent on the strength of the acid even for strong acids. In this way the order of decreasing strengths has been found to be $HClO_4 > HBr > H_2SO_4 > HCl > HNO_3$. Even the strongest known acid, perchloric acid, is a relatively poor conductor in acetic acid solution because the degree of ionization depends on the extent to which the solvent will accept protons from it to yield ions. Nevertheless, on account of the very marked tendency of the $CH_3CO_2H_2^+$ ion to lose the proton it has taken up from the acid, the solutions exhibit acidic properties, e.g., in catalysis, of an exceptionally strong nature; they have consequently been called **super-acid solutions** (J. B. Conant and N. F. Hall, 1927). At the other extreme, in a very strongly protophilic (basic) medium, such as liquid ammonia, it is probable that a weak acid like acetic would be highly ionized, and would behave like the stronger acids.

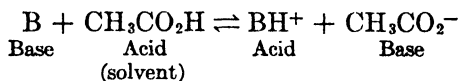
In a protogenic (acidic) solvent with no basic properties, such as hydrogen fluoride or hydrogen chloride, even the strongest acid is incapable of exhibiting acidic properties since there are no molecules available to accept protons. In fact many substances which are normally acids, but which have proton accepting properties to some extent, e.g., HNO_3 , behave as bases in hydrogen fluoride solution; the equilibrium established is



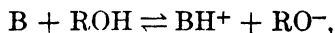
The weaker the acid HA in aqueous solution the more likely is it to be a proton acceptor and hence the more marked its basic properties in a proto-

genic medium. It is of interest to record that acetic acid is a relatively highly ionized base in hydrogen fluoride solution, whereas trichloroacetic acid is only a weak base (K. Fredenhagen, 1930). In an aprotic solvent, such as a hydrocarbon, neither acidic nor basic properties are evident; it is nevertheless possible for neutralization to occur, as will be seen later (p. 979).

Bases.—From the foregoing considerations it will be evident that the more acidic (protogenic) the solvent the more readily will the properties of a base become evident. For example, in acetic acid the equilibrium

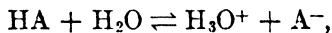


proceeds almost completely to the right even with weak bases; in fact all bases stronger than aniline appear equally strong in acetic acid solution (N. F. Hall, 1930). To distinguish between the strengths of bases it is necessary to use a medium which is only slightly protogenic in character; liquid ammonia or other base might be satisfactory. Water and the alcohols act as protogenic, as well as protophilic, solvents; such media are said to be **amphiprotic**. The base-solvent equilibrium is then



and since water has relatively strong protogenic properties moderately strong bases would exist in it largely as OH^- ions, and so their strengths could not be distinguished. It should be noted that the familiar strong bases, such as the alkali and alkaline-earth hydroxides, are invariably ionic in nature even in the solid state. The basic character of these strong bases is due to the OH^- ions which are always present, and interaction with the solvent is not an essential preliminary.

Dissociation Constants.—Consider an acid HA in water; the acid-base equilibrium is



and the equilibrium constant, the activity of the water being unity, is given by

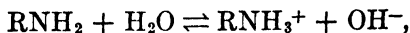
$$K_a = a_{\text{H}_3\text{O}^+} \times a_{\text{A}^-} / a_{\text{HA}}. \quad (187)$$

Since $a_{\text{H}_3\text{O}^+}$, the activity of the hydrogen ions, may be written a_{H^+} for simplicity, (187) becomes

$$K_a = a_{\text{H}^+} \times a_{\text{A}^-} / a_{\text{HA}}. \quad (188)$$

This is identical with the equation for the dissociation constant of an acid obtained by treating it in the conventional manner as a compound HA which ionizes to yield H^+ and A^- ions. The new point of view, therefore, introduces no fundamental change in the nature of the dissociation constant, and the results on p. 955 for acetic acid, and for other acids obtained in the same way, are still applicable.

When a weak base, e.g., ammonia or an amine RNH_2 , is dissolved in water, the equilibrium established is, according to the present view,



so that the dissociation constant is

$$K_b = a_{\text{RNH}_3^+} \times a_{\text{OH}^-} / a_{\text{RNH}_2}. \quad (189)$$

The classical method of considering the dissociation of a weak base was to write the equilibrium as

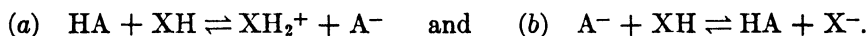


i.e.,

$$K'_b = a_{\text{RNH}_3^+} \times a_{\text{OH}^-} / a_{\text{RNH}_3\text{OH}}. \quad (190)$$

In deriving K'_b , however, it was generally assumed that $a_{\text{RNH}_3\text{OH}}$ was equal to the activity (concentration) of the *whole* of the amine which was not ionized; in other words, it was equivalent to a_{RNH_2} of (189) so that the constants K'_b and K_b are identical. Once again, therefore, the proton transfer theory of acids and bases requires no change to be made in the value of the dissociation constant; the data given in the older literature can consequently still be employed.

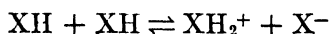
In an amphiprotic solvent, such as water or an alcohol, the conjugate acid and base can function simultaneously; it is of interest to consider their relative dissociation constants in a given medium. If XH represents an amphiprotic molecule, then it can take up a proton to give XH_2^+ or it can yield a proton to form X^- . The equilibria between the solvent and (a) the acid HA , and (b) its conjugate base A^- , may be written



The dissociation constants of acid and conjugate base in the same solvent are then

$$K_a = a_{\text{XH}_2^+} \times a_{\text{A}^-} / a_{\text{HA}} \quad \text{and} \quad K_b = a_{\text{HA}} \times a_{\text{X}^-} / a_{\text{A}^-} \quad (191)$$

where XH_2^+ is the hydrogen ion in the given medium and X^- is the solvent anion. Since the solvent is amphiprotic it can itself function as an acid and a base; hence the equilibrium



must be established, and assuming the activity of the solvent to be unity, as will be the case for dilute solutions, then

$$K_I = a_{\text{XH}_2^+} \times a_{\text{X}^-}. \quad (192)$$

The quantity K_I , the product of the activities of the acidic and basic ions formed by the solvent, is called the **ionic product of the solvent**. Combination of (191) and (192) gives

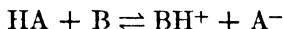
$$K_a K_b = K_I, \quad (193)$$

so that the dissociation constant of an acid is inversely proportional to that of its conjugate base, and *vice versa*, in the given amphiprotic medium.

Strengths of Acids and Bases.—The strength of an acid should be governed by its tendency to give up protons, and a measure of this would be the equilibrium constant of the process $\text{HA} \rightleftharpoons \text{p}^+ + \text{A}^-$, where p^+ is the proton; that is,

$$K_{\text{HA}} = a_{\text{p}^+} \times a_{\text{A}^-} / a_{\text{HA}}, \quad (194)$$

where a_{p^+} is the activity of *protons*. This quantity is the *fundamental* dissociation constant of the acid and, unlike the dissociation constant usually measured, it is independent of the nature of the solvent. Unfortunately it is not possible to determine K_{HA} directly, but relative values for a series of acids may be derived from observations in an aprotic medium. In such a solvent neither acid nor base can function alone, but if a base B is added to an acid HA, the proton from the latter can pass to the former, so that the neutralization equilibrium



is established, and application of the law of equilibrium gives

$$K = a_{\text{BH}^+} \times a_{\text{A}^-} / a_{\text{HA}} \times a_{\text{B}}. \quad (195)$$

The fundamental dissociation constants of the acids HA and BH^+ are given by $K_{\text{HA}} = a_{\text{p}^+} \times a_{\text{A}^-} / a_{\text{HA}}$ and $K_{\text{BH}^+} = a_{\text{p}^+} \times a_{\text{B}} / a_{\text{BH}^+}$, so that it follows from (195) that

$$K = K_{\text{HA}} / K_{\text{BH}^+}. \quad (196)$$

The equilibrium constant of the neutralization reaction in an aprotic solvent is thus a measure of the relative strengths of the neutralized acid and of the conjugate acid corresponding to the base used. The value of K can be ascertained in some instances if the color of the base B is different from that of its conjugate acid; if equivalent amounts of HA and B are mixed, the color of the resulting solution then gives the ratio $c_{\text{BH}^+} / c_{\text{B}}$, and hence also $c_{\text{A}^-} / c_{\text{HA}}$. Assuming the concentration ratios to be equal to the activity ratios it is possible to evaluate K . In this manner the strengths of BH^+ and HA may be compared, but it is more usual to carry out a series of measurements with one acid HA and various bases B; the latter, or their conjugate acids, can then be arranged in order of increasing strengths. Similarly, a comparison of various acids could be made by using the same base B in each case.

The procedure described above is limited to suitable colored substances, and so it is more usual to compare the strengths of a series of acids or bases in a given solvent; since the solvent merely acts as a means of removing or supplying protons the *relative* strengths should be independent of its nature, and this appears to be approximately true. Since the hydrogen ion, e.g., H_3O^+ , in water, is an acid, its fundamental dissociation constant $K_{\text{H}_3\text{O}^+}$ is equal to $a_{\text{p}^+} \times a_{\text{H}_2\text{O}} / a_{\text{H}_3\text{O}^+}$; combination of this

result with (187) and (194) gives

$$K_a = (K_{HA}/K_{H^+O})a_{H_2O}. \quad (197)$$

The activity of the water can be taken as unity, and $K_{H_3O^+}$ for H_3O^+ is a constant; hence the conventional dissociation constant K_a is directly proportional to the fundamental constant K_{HA} which measures the true strength of the acid. According to the foregoing arguments the values of K_a for a series of acids should be in the same order irrespective of the nature of the solvent. This is, however, only approximately true, for the electrical forces between the ions are affected differently by the possible orientation of solvent molecules in the immediate vicinity of the ions. Nevertheless, the order of the dissociation constants in a particular solvent, e.g., water, gives an approximate indication of the order of the strengths of a series of related acids. The strengths of bases may be compared in a similar manner by their normal dissociation constants in a given solvent.²⁸

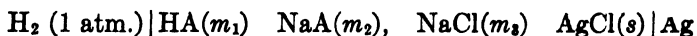
Determination of Dissociation Constants.—The method for obtaining dissociation constants from conductance measurements has been already described. If accurate results are desired it is necessary to calculate the degree of dissociation correctly, and to allow for deviation of the activity coefficients from unity. Another common procedure is to measure the hydrogen ion concentration in a solution containing a known amount of the acid and its salt with a strong base. For an acid HA the dissociation constant K_a is given by

$$K_a = \frac{a_{H^+} \times a_{A^-}}{a_{HA}} = \frac{c_{H^+} \times c_{A^-}}{c_{HA}} \cdot \frac{f_{H^+} \times f_{A^-}}{f_{HA}}, \quad (198)$$

where H^+ represents the hydrogen ion as it exists in solution. In a mixture of a moles of a weak acid and b moles of its highly ionized salt, the latter may be regarded as completely dissociated into A^- ions, so that $c_{A^-} = b$; further, the excess of anions will, according to the law of equilibrium, repress the ionization of the acid HA, so that the latter may be assumed to be completely nonionized and hence $c_{HA} = a$.^{*} By means of a suitable form of hydrogen electrode (p. 994), a quantity which is presumed to be a_{H^+} is obtained, and consequently all the terms in (198), except f_{A^-} and f_{HA} , are available; the latter may be taken as unity (p. 970), and the former can be estimated or derived from the Debye-Hückel equations, and so K_a can be determined. The procedure should be repeated for a number of mixtures so that the results may be confirmed. In the special case of a mixture of equal amounts of acid and its salt, or if the acid is half-neutralized, then $c_{HA} = c_{A^-}$; the hydrogen ion activity is then approximately equal to the dissociation constant. The dissociation constant of a weak base may be obtained in an analogous manner from mixtures of the base and its salt with a strong acid.

* These assumptions are only justifiable if the acid is fairly weak; for more accurate equations see p. 1003.

A more exact method, involving E.M.F. measurements, has been described by H. S. Harned (1930 *et seq.*). The cell



is a form of hydrogen-silver chloride cell, and its E.M.F. is represented by [cf. equation (79)]

$$E = E^0 - \frac{RT}{F} \ln a_{\text{H}^+} a_{\text{Cl}^-} = E^0 - \frac{RT}{F} \ln (m_{\text{H}^+} m_{\text{Cl}^-}) (\gamma_{\text{H}^+} \gamma_{\text{Cl}^-}), \quad (199)$$

where E^0 is known to be +0.2224 volt at 25° c. (p. 943); the activities and activity coefficients refer, of course, to the particular values in the given cell. If the definition of K_a as given by (198), using molalities, is introduced, it follows that

$$E - E^0 + \frac{RT}{F} \ln \frac{m_{\text{HA}} m_{\text{Cl}^-}}{m_{\text{A}^-}} = - \frac{RT}{F} \ln \frac{\gamma_{\text{Cl}^-} \gamma_{\text{HA}}}{\gamma_{\text{A}^-}} - \frac{RT}{F} \ln K_a. \quad (200)$$

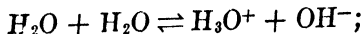
If E is measured, and E^0 is known, the left-hand side can be evaluated for various concentrations of acid (m_1), salt (m_2) and sodium chloride (m_3); m_{Cl^-} may be put equal to m_3 ; m_{HA} is given by $m_1 - m_{\text{H}^+}$; and m_{A^-} is $m_2 + m_{\text{H}^+}$, a sufficiently accurate value of m_{H^+} being obtained from an approximate dissociation constant. The resulting quantities are plotted against the ionic strength of the solution in the cell, and extrapolated to zero concentration; the activity coefficient fraction on the right-hand side of (200) becomes unity and its logarithm zero, so that the intercept gives $-(RT/F) \ln K_a$. This is probably the most accurate method for determining dissociation constants. It could be adapted to the study of bases, but the majority of these are too volatile or else too weak for accurate measurements to be of any value. It has, however, been used for amino-acids.

The dissociation constants, and hence the strengths, of acids and bases can often be compared by methods depending on the determination of their catalytic coefficients in certain reactions; this subject will be discussed in Chapter XIII. Some dissociation constants of a number of acids and bases in aqueous solution at 25° c. are quoted in Table 163.

TABLE 163. DISSOCIATION CONSTANTS OF ACIDS AND BASES AT 25° C.

Acids		Bases	
HCN	7.24×10^{-10}	NH ₃	1.74×10^{-5}
HCO ₂ H	1.77×10^{-4}	CH ₃ NH ₂	5.01×10^{-4}
CH ₃ CO ₂ H	1.75×10^{-5}	(CH ₃) ₂ NH	7.41×10^{-4}
C ₂ H ₅ CO ₂ H	1.34×10^{-5}	(CH ₃) ₃ N	7.40×10^{-5}
CH ₃ ClCO ₂ H	1.35×10^{-3}	C ₂ H ₅ NH ₂	5.62×10^{-4}
C ₆ H ₅ CO ₂ H	6.31×10^{-5}	C ₆ H ₅ NH ₂	4.07×10^{-10}
C ₆ H ₅ OH	1.20×10^{-10}	C ₆ H ₅ CH ₂ NH ₂	2.29×10^{-5}

Ionic Equilibria in Water.—Since water is an amphiprotic substance there must exist in pure water and in all aqueous solutions the equilibrium



hence, assuming the activity of the water to be constant, viz., unity, the ionic product is

$$K_w = a_{H_3O^+} \times a_{OH^-} = (c_{H^+} c_{OH^-})(f_{H^+} f_{OH^-}), \quad (201)$$

where c_{H^+} and f_{H^+} represent the concentration and activity coefficient, respectively, of the hydrogen ion H_3O^+ . Provided the ionic strength of the medium is low, the activity coefficients are approximately unity,* so that (201) becomes

$$k_w = c_{H^+} \times c_{OH^-}, \quad (202)$$

which must be regarded as an approximate relationship applicable in dilute solutions only. In an exactly neutral solution, the hydrogen and hydroxyl ion concentrations are equal, i.e., $c_{H^+} = c_{OH^-}$, and hence from (202), each of these is equal to $\sqrt{k_w}$. At ordinary temperatures the ionic product can be assumed to be 10^{-14} , and so in a neutral solution the hydrogen and hydroxyl ion concentrations are both 10^{-7} g. ion per liter. If c_{H^+} exceeds this value the solution is said to be acid; in such a solution c_{OH^-} must be less than 10^{-7} g. ion per liter. On the other hand, if c_{H^+} is less than 10^{-7} and c_{OH^-} greater, then the aqueous solution is alkaline; whether the solution is acid, alkaline or neutral, however, the product $c_{H^+} \times c_{OH^-}$ must be 10^{-14} at room temperature.

Since hydrogen ion concentrations can normally vary from about 1 g. ion per liter in a N solution of strong acid to 10^{-14} in a solution of alkali, a more convenient scale for expressing such concentrations is desirable; the one suggested by S. P. L. Sørensen (1909) has come into general use. The *negative* logarithm of the hydrogen ion activity is called the pH , or p_H , of the solution; thus,

$$pH = -\log a_{H^+} \quad \text{or} \quad a_{H^+} = 10^{-pH}. \quad (203)^*$$

The logarithmic, or exponential, method has been extended to other instances, e.g., to represent hydroxyl ion activities pOH is used; thus,

$$pOH = -\log a_{OH^-}, \quad (204)$$

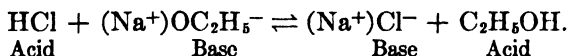
and dissociation constants are described by pK or pk , which are equal to $-\log K$ or $-\log k$, respectively. It is consequently possible to write (202) in the form

$$pH + pOH = pK_w = 14. \quad (205)$$

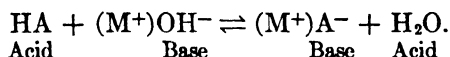
In a neutral solution at ordinary temperatures the pH is 7, but in an acid solution it is *less*, and in an alkaline medium *greater*, than this value.

* Because of the uncertainties relating to single ion activities, pH and pOH are often defined, approximately, in terms of concentrations in place of the respective activities.

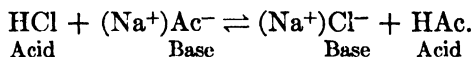
Neutralization and Hydrolysis.—In its widest sense the term neutralization refers to the reaction between equivalent amounts of acid and base, but the products are not necessarily to be described as a salt and a molecule of solvent, as in the conventional description of neutralization; they are actually the conjugate base and acid of the acid and base, respectively, used in the neutralization. For example, the acid may be HCl and the base NaOC_2H_5 , of which the OC_2H_5^- ion is the real base; the neutralization is then



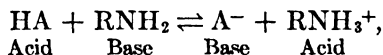
If the acid is written HA and the base is of the conventional type, e.g., MOH , then the neutralization is



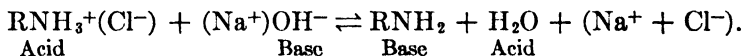
It is seen that the anion of every acid is to be regarded as a base: for example, the acetate ion is a base, and the displacement reaction between sodium acetate and a strong acid, e.g., HCl , is really a neutralization; thus,



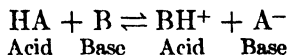
When an anhydro-base, e.g., ammonia or an amine, is neutralized by an acid HA , the reaction is



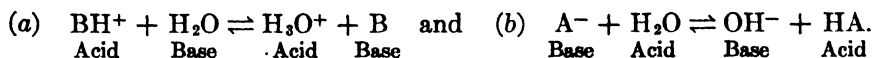
so that the ammonium ion, RNH_3^+ , is an acid, being the conjugate acid to the base RNH_2 . The displacement reaction between an ammonium, or substituted ammonium, salt and a strong base, e.g., NaOH , is therefore a neutralization in its widest sense; thus,



In general, a neutralization can be written



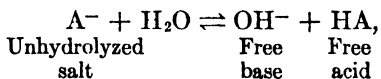
and the extent to which the reaction proceeds depends not only on the relative strengths of the two acids concerned, but also on the nature of the solvent. If it is, like water, an amphiprotic solvent, which is capable of acting as an acid or a base, it may interact with the neutralization products in two ways, viz.,



In (a) the original base is regenerated, whereas in (b) the acid is produced, and so these reactions, between water and the acid and base conjugate to those taking part in the neutralization, result in a partial reversal of neutralization. The general name for this phenomenon is **solvolysis** or **lyolysis**; when the solvent is water it is known as **hydrolysis**. If the solvent is exclusively protophilic (basic) reaction (a) only will occur, but if it is exclusively protogenic (acidic) then only (b) will be possible.

Hydrolysis and Strength of Acid and Base.—According to (193) the strength of an acid or base is inversely proportional to that of its conjugate form; hence if HA is a strong acid, e.g., HCl, its conjugate base, i.e., Cl⁻, will be extremely weak, and hydrolysis (b) occurs to a negligible extent. If the base used for neutralization, e.g., NaOH, is also strong, its conjugate form, i.e., H₂O, will be a very weak acid, and consequently hydrolysis of type (a) will take place only to a minute degree. It is evident, therefore, that neutralization between a strong acid and a strong base may be assumed to go to completion, there being no hydrolysis.* When the salt of a strong base and a strong acid, e.g., NaCl, is dissolved in water, the ions produced, viz., Na⁺ and Cl⁻, have no detectable tendency to interact with the solvent, and the solution remains neutral. If acid or base is weak, however, there is appreciable hydrolysis, and three cases may be considered involving (i) weak acid, (ii) weak base, and (iii) weak acid and base.

(i) **Weak Acid and Strong Base.**—If the acid HA is weak, the conjugate base A⁻ will be fairly strong, and interaction with the solvent, acting as an acid, will take place to a definite extent; hence the addition of 1 equiv. of a strong base to 1 equiv. of a weak acid does not result in complete neutralization. It follows, also, that when the salt of a weak acid and strong base, e.g., NaA, is dissolved in water the A⁻ ions produced on dissociation will establish an equilibrium with water molecules



resulting in the partial re-formation of the free weak acid (HA) and the strong base (OH⁻) from which the salt was constituted. As a consequence of the weakness of HA, and the appreciable strength, therefore, of the base A⁻, hydrolysis—that is to say, a partial reversal of neutralization—occurs. The production of free OH⁻ ions must be accompanied by a decrease of hydrogen ions, and so the result of hydrolysis is that the solution of the salt of a weak acid and strong base reacts alkaline.

Applying the law of equilibrium to the hydrolytic reaction, the activity of the water being taken as unity, the result is

$$K_A = \frac{a_{\text{OH}^-} \times a_{\text{HA}}}{a_{\text{A}^-}} = \frac{c_{\text{OH}^-} \times c_{\text{HA}}}{c_{\text{A}^-}} \cdot \frac{f_{\text{OH}^-} \times f_{\text{HA}}}{f_{\text{A}^-}}, \quad (206)$$

* This conclusion would have to be modified if the solvent were *very strongly* protogenic, e.g., liquid HF.

the quantity K_h being called the **hydrolysis constant** of the salt. In solutions of low ionic strength the activity coefficient f_{HA} of the undissociated molecules of acid may be taken as unity, and further the ratio of the coefficients for two univalent ions, i.e., f_{OH^-}/f_{A^-} , is also unity by the Debye-Hückel relationship (131); equation (206) then reduces to the less exact form

$$k_h = \frac{c_{OH^-} \times c_{HA}}{c_{A^-}}, \quad (207)$$

applicable to dilute solutions. If a solution of the salt NaA is at a concentration of c moles per liter and x is the degree of hydrolysis, i.e., the fraction of each mole of salt hydrolyzed when equilibrium is attained, then $c(1 - x)$ moles per liter is the concentration of unhydrolyzed salt. Since this is almost completely dissociated into A^- ions, it is possible to write

$$c_{A^-} = c(1 - x).$$

The concentrations of HA and of OH^- produced by hydrolysis must be equal to cx , so that

$$c_{HA} = c_{OH^-} = cx,$$

and substitution in (207) gives

$$k_h = cx^2/(1 - x); \quad (208)$$

$$\therefore x = -\frac{k_h}{2c} + \sqrt{\frac{k_h^2}{4c^2} + \frac{k_h}{c}}. \quad (209)$$

From (209) it is possible to calculate the degree of hydrolysis at any desired concentration, provided the hydrolysis constant is known. If x is small, then (208) reduces to $k_h \approx cx^2$, and hence

$$x \approx \sqrt{k_h/c}, \quad (210)$$

so that the degree of hydrolysis of a given salt is roughly proportional to the square root of the hydrolysis constant and inversely proportional to the square root of the concentration.

Since the relationships $c_{H^+} \times c_{OH^-} = k_w$ for water, and $c_{H^+} \times c_{A^-}/c_{HA} = k_a$ for the weak acid must hold in the same solution as the hydrolytic equation (207), it follows that

$$k_h = k_w/k_a \quad \text{or} \quad K_h = K_w/K_a, \quad (211)$$

and so the hydrolytic constant is related to the ionic product of water and to the strength of the acid; this is in agreement with general expectations (p. 978). If k_h as given by (211) is substituted in (210), it is seen

$$x = \sqrt{k_w/k_a c}. \quad (212)$$

The figures in Table 164 give the values of the degree of hydrolysis at 25° c. calculated from (212), for the salts of a series of acids of k_a varying

TABLE 164. DEGREE OF HYDROLYSIS OF SALT OF WEAK ACID AND STRONG BASE AT 25° C.

k_a	k_b	Concentration of Salt Solution			
		0.001 N	0.01 N	0.1 N	N
10^{-4}	10^{-10}	3.3×10^{-4}	10^{-4}	3.2×10^{-5}	10^{-5}
10^{-6}	10^{-8}	3.2×10^{-3}	10^{-3}	3.2×10^{-4}	10^{-4}
10^{-8}	10^{-6}	3.2×10^{-2}	10^{-2}	3.2×10^{-3}	10^{-3}
10^{-10}	10^{-4}	0.27	0.095	3.2×10^{-2}	10^{-2}

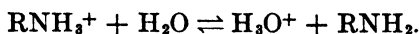
from 10^{-4} to 10^{-10} , at concentrations from 0.001 N to N. The degree of hydrolysis at any concentration is greater the weaker the acid constituting the salt and the more dilute the solution. In a 0.001 N solution of the sodium salt of an acid of k_a equal to 10^{-10} , e.g., a phenol, the salt is hydrolyzed to the extent of 27 per cent.

It has been seen that $c_{OH^-} = cx$, and since $c_{H^+} \times c_{OH^-} = k_w$, it follows that $c_{H^+} = k_w/cx$, and introducing the value of x from (212),

$$c_{H^+} = \sqrt{k_w k_a / c}; \quad \therefore pH = \frac{1}{2}pk_w + \frac{1}{2}pk_a + \frac{1}{2} \log c. \quad (213)$$

It follows, therefore, that the pH, or alkalinity, of the solution increases with increasing concentration of the salt, although the degree of hydrolysis decreases.

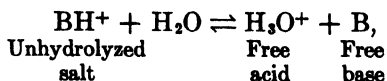
(ii) **Weak Base and Strong Acid.**—When the base B is weak, the conjugate acid BH^+ will have appreciable strength and will set up an equilibrium, reaction (b), p. 978, with the solvent acting as a base. It follows, therefore, that if a salt of a strong acid and weak base is dissolved in water there will be a partial reversal of neutralization; that is, hydrolysis will occur. If the weak base is of the type RNH_2 the conjugate acid is RNH_3^+ , and when the salt, e.g., RNH_3Cl , is dissolved in water it dissociates to give RNH_3^+ which establishes the hydrolytic equilibrium



When the weak base is a metallic hydroxide, it is probable that the conjugate acid is the hydrated ion of the metal, e.g., $Fe(H_2O)_6^{+++}$, $Cu(H_2O)_4^{++}$, resulting from the ionization of the salt; the hydrolysis must then be represented by



where $M(H_2O)_{n-1}OH$ is the weak base. Writing the hydrolytic equilibrium in the general form,



the equilibrium constant is given by

$$K_A = \frac{a_{H^+} \times a_B}{a_{BH^+}} = \frac{c_{H^+} \times c_B}{c_{BH^+}} \cdot \frac{f_{H^+} \times f_B}{f_{BH^+}}. \quad (214)$$

Making the same assumptions as previously with regard to the activity coefficients, the approximate equation, for dilute solution, is

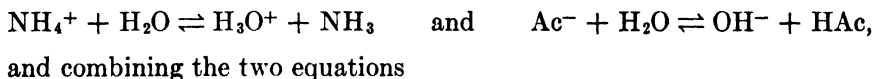
$$k_h = c_{H^+} \times c_B / c_{BH^+} = k_w / k_b, \quad (215)$$

where the dissociation constant k_b of the base is $c_{BH^+} \times c_{OH^-} / c_B$. By the use of treatment similar to that described for the salt of a weak acid and strong base, it is possible to derive equations analogous to (208), (209) and (210). The corresponding form of (212) is $x = \sqrt{k_w / k_b c}$, and since c_{H^+} is cx , it follows that

$$c_{H^+} = \sqrt{k_w c / k_b}; \quad \therefore pH = \frac{1}{2}pk_w - \frac{1}{2}pk_b - \frac{1}{2}\log c. \quad (216)$$

The general rule that the degree of hydrolysis increases with dilution and with decreasing strength of the base, holds for salts of strong acids and weak bases. It is evident from (216) that the pH of the solution must be less than $\frac{1}{2}pk_w$, i.e., less than 7; that is to say, solutions of salts of the type under consideration have an acid reaction. This conclusion can also be reached from the equation of the hydrolytic equilibrium; the formation of the H_3O^+ ions shows that the solution must be acid.

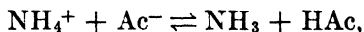
(iii) **Weak Acid and Weak Base.**—If acid and base from which a salt is constituted are both weak, the respective conjugate base and acid have appreciable strength and so both interact with the solvent. When a salt such as ammonium acetate is dissolved in water the ions, NH_4^+ and Ac^- , acting as acid and base, respectively, take part in the equilibria



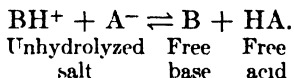
or, in general,



Since the normal equilibrium between water molecules and hydrogen and hydroxyl ions exists in any case, this may be subtracted from the hydrolytic equilibrium, so that the result may be represented by



or, in general,



The expression for the equilibrium (hydrolysis) constant is then

$$K_h = \frac{a_B \times a_{HA}}{a_{BH^+} \times a_{A^-}} = \frac{c_B \times c_{HA}}{c_{BH^+} \times c_{A^-}} \cdot \frac{f_B \times f_{HA}}{f_{BH^+} \times f_{A^-}}. \quad (217)$$

The activity coefficients f_B and f_{HA} may be assumed to be unity, and

if the solution is dilute f_{BH^+} and f_{A^-} do not differ greatly from unity, so that the approximate form of (217) is

$$k_a = \frac{c_{\text{B}} \times c_{\text{HA}}}{c_{\text{BH}^+} \times c_{\text{A}^-}} = \frac{k_w}{k_a k_b}, \quad (218)$$

where the dissociation constants, k_a and k_b , of the weak acid and base constituting the salt are defined in the usual manner.

Suppose the original concentration of the salt is c moles per liter, and x is the fraction hydrolyzed at equilibrium; then c_{BH^+} and c_{A^-} may be put equal to the concentration of unhydrolyzed salt, i.e., $c(1-x)$, which is assumed to be completely dissociated. The concentrations of free acid and free base, i.e., c_{B} and c_{HA} , may be taken as equal to cx , so that by (218),

$$k_a = x^2/(1-x)^2 \quad \text{and} \quad x = \sqrt{k_a}/(1 + \sqrt{k_a}). \quad (219)$$

If k_a is small in comparison with unity, then

$$x \approx \sqrt{k_a} = \sqrt{k_w/k_a k_b}. \quad (220)$$

The hydrogen ion concentration of the hydrolyzed salt solution may be calculated by using the expression for the dissociation constant $k_a = c_{\text{H}^+} \times c_{\text{A}^-}/c_{\text{HA}}$,

$$\therefore c_{\text{H}^+} = k_a c_{\text{HA}}/c_{\text{A}^-} = k_a cx/c(1-x) = k_a x/(1-x). \quad (221)$$

By (219), $x/(1-x)$ is equal to $\sqrt{k_a}$, i.e., $\sqrt{k_w/k_a k_b}$, and so

$$c_{\text{H}^+} = \sqrt{k_w k_a/k_b} \quad \text{or} \quad \text{pH} = \frac{1}{2}pk_w + \frac{1}{2}pk_a - \frac{1}{2}pk_b. \quad (222)$$

When $k_a = k_b$, that is to say, the dissociation constants of acid and base are equal, the pH is $\frac{1}{2}pk_w$, so that the solution will be neutral, *in spite of hydrolysis*. If $k_a > k_b$, then $pk_a < pk_b$ and the salt solution will be acid, but if $k_a < k_b$ the solution will be alkaline.

It appears from (220) and (222) that the extent of hydrolysis and the hydrogen ion concentration are independent of the concentration of the salt; this is strictly true only if $c_{\text{BH}^+} = c_{\text{A}^-}$ and $c_{\text{B}} = c_{\text{HA}}$, as assumed above. This condition will be realized if k_a and k_b are equal, but not otherwise. It is important to remember that if the acid BH^+ and the base A^- have different strengths they will interact with water to different extents, so that at equilibrium c_{BH^+} and c_{A^-} are not necessarily equal; similarly, the products B and HA may not be at the same concentration. -It is possible, however, to write

$$c = c_{\text{BH}^+} + c_{\text{B}} = c_{\text{HA}} + c_{\text{A}^-}, \quad (223)$$

where c is the total concentration of salt, and (223) together with the condition of electrical neutrality, i.e.,

$$c_{\text{H}^+} + c_{\text{BH}^+} = c_{\text{OH}^-} + c_{\text{A}^-}, \quad (224)$$

and the usual formulae for k_a , k_b and k_w permit a more exact relationship for c_{H^+} to be derived. The difference between this result and (222) is only appreciable for dilute solutions, even when k_a and k_b differ by several powers of 10.

Determination of Hydrolysis Constant.—(i) *Dissociation Constant Method:* According to (211), (215) and (218) the hydrolysis constant is related to k_w and either k_a or k_b , for a salt of one-sided weakness, or to both k_a and k_b if both acid and base are weak. The methods for determining dissociation constants have been already described, and hence only k_w is required; the means whereby this may be obtained will be considered shortly. Once k_a is known it is possible to calculate the degree of hydrolysis at any desired concentration.

(ii) *Hydrogen Ion Method.*—If the acid constituting the salt is weak, then c_{OH^-} is equal to cx , and hence $c_{H^+} = k_w/cx$; on the other hand, for a salt of a weak base, $c_{H^+} = cx$, as seen above. By determining the hydrogen, or hydroxyl, ion concentration of the salt solution of known total concentration it is therefore possible to evaluate the degree of hydrolysis x , and hence the hydrolysis constant. The most reliable method for obtaining the hydrogen ion concentration is by the use of some form of hydrogen electrode (p. 994), or by means of an indicator (p. 1001). In the earlier studies of hydrolysis catalytic methods were frequently employed, e.g., hydrolysis of methyl acetate, decomposition of nitrosotriacetoneamine, etc., but it is now known that it is essential to choose reactions which are catalyzed exclusively, or almost so, by hydrogen or by hydroxyl ions (p. 1133). If both acid and base are weak then c_{H^+} alone is insufficient to give k_b ; it follows from (219) and (221) that c_{H^+} is equal to $k_a\sqrt{k_b}$, and so it is necessary to know k_a in addition to c_{H^+} for k_b to be evaluated.

(iii) *Conductance Method.*—In a solution of 1 g. equiv. per liter of a salt of a strong acid and weak base, of which a fraction x is hydrolyzed, there will be present $1 - x$ g. equiv. of unhydrolyzed salt, and x g. equiv. of both free acid and base. If the base is very weak, it may be regarded as completely non-ionized, and so it will contribute nothing toward the total conductance of the solution. The conducting power of 1 g. equiv. of a hydrolyzable salt is thus made up of the conductance of $1 - x$ g. equiv. of unhydrolyzed salt, and x g. equiv. of free acid, i.e.,

$$\Lambda = (1 - x)\Lambda_c + x\Lambda_{HA}, \quad (225)$$

where Λ is the apparent equivalent conductance of the solution, i.e., $1000 \times$ measured specific conductance/conc. in g. equiv. per liter (see p. 889), Λ_c is the equivalent conductance of the *unhydrolyzed* salt at the particular concentration used in the experiment, and Λ_{HA} is the equivalent conductance of the free acid. It follows from (225) that

$$x = (\Lambda - \Lambda_c)/(\Lambda_{HA} - \Lambda_c). \quad (226)$$

In order to determine x it is necessary to know Λ , Λ_c and Λ_{HA} , of which Λ is obtained by direct measurement of the salt solution, as already indicated. The value of Λ_{HA} is generally taken as the equivalent conductance of the strong acid at infinite dilution, since its concentration is very small; it is probably more accurate, however, to use the equivalent conductance at the same total ionic strength as exists in the salt solution. The determination of Λ_c for the unhydrolyzed salt, presents a special problem, which was solved by G. Bredig

(1894). It follows from the law of equilibrium that the addition of free base must repress the hydrolysis of the salt solution, and if the former is a very poor conductor, the conductance of a mixture of salt and excess free base is virtually that of the unhydrolyzed salt. For example, the apparent equivalent conductance of $N/1024$ aniline hydrochloride, i.e., Λ , is 144.0 at 18°C .; if this solution is made $N/64$ with respect to free added aniline, the equivalent conductance is reduced to 103.3, and the addition of more aniline results in no further reduction. It may be assumed, therefore, neglecting the conductance of the free aniline, that the equivalent conductance of the unhydrolyzed salt, i.e., Λ_s , is 103.3 for a $N/1024$ solution. Taking Λ_{HCl} as 380.0, it follows from (226) that x is 0.147, and hence k_a for aniline hydrochloride at 18°C . is 2.5×10^{-5} . The method described can be readily adapted to the study of salts of weak acids; in each case the chief source of error is the neglect of the conductance of the added weak base or weak acid, although it is possible to apply corrections if necessary.

(iv) *Distribution Method*.—This method depends on the principles described on p. 740; it can be used if one constituent of the salt, but not the salt itself or the other constituent, is soluble in a liquid immiscible with water. Consider the salt of a weak base, e.g., aniline hydrochloride; the free base is soluble in benzene in which it has a normal molecular weight, and suppose its distribution coefficient between water and benzene is known. A definite volume of an aqueous solution of the salt at known concentration is shaken with a given volume of benzene, and the amount of free aniline in the latter is determined by analysis; from the distribution ratio the concentration of free base (c_b) in the aqueous solution can be readily calculated. The amount of free strong acid in the water layer is equivalent to the free base in the water and in the benzene, and since both the latter are known, the former and the corresponding concentration (c_{H^+}) can be calculated. In order to calculate the hydrolysis constant by (215) it is necessary to know in addition the concentration of unhydrolyzed salt (c_{BH^+}) in the water; this is equal to the initial concentration less that of the free acid.²⁹

Determination of the Ionic Product of Water.—(i) *Salt Hydrolysis Method*: A reversal of the procedure described under method (i) on p. 989 permits the ionic product of water to be determined if the hydrolysis constant of a salt and the dissociation constant of its constituent weak acid, or weak base, are known. For example, k_a for aniline hydrochloride was found to be 2.5×10^{-5} at 18°C . by method (iii) above; since k_b for aniline is 4×10^{-10} , it follows that $k_w = k_b k_a = 10^{-14}$ at 18° .

(ii) *Conductance Method*.—The conductivity of perfectly pure water will be due to the hydrogen and hydroxyl ions in equilibrium with water molecules. The purest water known was obtained by F. W. Kohlrausch and A. Heydweiller (1894), after forty-eight distillations under reduced pressure, and its specific conductance was $0.043 \times 10^{-6} \text{ ohm}^{-1} \text{ cm}^{-1}$ at 18°C . Even this water probably contained traces of impurities, and it was estimated that the conductance of a 1 cm. cube of pure water would be $0.0384 \times 10^{-6} \text{ ohm}^{-1} \text{ cm}^{-1}$. The ionic conductances of H^+ and OH^- ions at very small concentrations, such as exist in water, are 315.2 and 173.8, respectively at 18° , and so the total conductance of 1 g. equiv. of H^+ and OH^- ions would be 489 $\text{ohms}^{-1} \text{ cm}^2$. It follows, therefore, that 1 cc. of water contains $0.0384 \times 10^{-6}/489$, i.e., 0.78×10^{-10} g. equiv. of hydrogen and hydroxyl ions. The concentrations in g. ion per liter

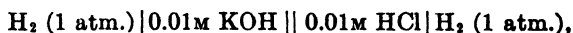
are thus 0.78×10^{-7} , and hence at 18°C .

$$k_w = c_{\text{H}^+} \times c_{\text{OH}^-} = (0.78 \times 10^{-7})^2 = 0.61 \times 10^{-14}.$$

The ionic products of ethyl alcohol, acetic acid and other solvents have been obtained in a similar manner.

(iii) *Reaction Velocity Method*.—This method depends ultimately on the determination of the hydrogen and hydroxyl ion concentration in pure water by their catalytic action; it does not, however, give results of any great accuracy. The best form is that used by H. M. Dawson (1927), depending on the determination of the hydrogen ion concentration at the catalytic minimum. If the catalytic coefficients of hydrogen and hydroxyl ions are known, the value of k_w can be obtained by (169), p. 1135.

(iv) *E.M.F. Methods*.—The most reliable values for the ionic (activity) product of water are derived from measurements of suitable cells. For example, the E.M.F. of the cell



free from liquid junction potential, is given by

$$E = \frac{RT}{F} \ln \frac{a_{\text{H}^+}}{a_{\text{H}^+}''} = \frac{RT}{F} \ln \frac{a_{\text{H}^+}' \times a_{\text{OH}^-}''}{K_w}, \quad (227)$$

where a_{H^+}' and a_{OH^-}'' are the activities of hydrogen and hydroxyl ions in the acid and alkaline solutions, respectively. The E.M.F. of the cell was not measured directly, but the potential of each half-cell was determined by combination with a calomel electrode, an allowance being made for the liquid junction potential. The values found (G. N. Lewis, *et al.*, 1917) were -0.4030 and -0.9904 , so that the E.M.F. of the cell is 0.5874 at 25°C . If the activity coefficients of hydrogen and hydroxyl ions in 0.01M strong acid and alkali are assumed to be 0.930 , then a_{H^+}' and a_{OH^-}'' are each 0.0093 , and substitution in (227) gives K_w as 1.01×10^{-14} at 25° .

The difficulty of liquid junctions can be avoided by the method of E. J. Roberts (1930). The E.M.F. of the cell



which is exactly analogous to that used for the dissociation constants of acids (p. 981), is

$$E = E^0 - \frac{RT}{F} \ln a_{\text{H}^+} a_{\text{Cl}^-} = E^0 - \frac{RT}{F} \ln K_w a_{\text{Cl}^-} / a_{\text{OH}^-}, \quad (228)$$

where E^0 is, as before, 0.2224 at 25° . Substituting $m_{\text{Cl}^-} \gamma_{\text{Cl}^-}$ for a_{Cl^-} and $m_{\text{OH}^-} \gamma_{\text{OH}^-}$ for a_{OH^-} , and rearranging, (228) becomes

$$E - E^0 + \frac{RT}{F} \ln \frac{m_{\text{Cl}^-}}{m_{\text{OH}^-}} = - \frac{RT}{F} \ln \frac{\gamma_{\text{Cl}^-}}{\gamma_{\text{OH}^-}} - \frac{RT}{F} \ln K_w. \quad (229)$$

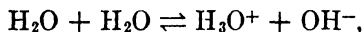
It is clear, therefore, that if the left-hand side of this equation, where E is the measured E.M.F. of the cell, and m_{Cl^-} and m_{OH^-} are equal to m_2 and m_1 , respectively, is plotted against the ionic strength and extrapolated to infinite dilution, when $\ln \gamma_{\text{Cl}^-} / \gamma_{\text{OH}^-}$ is zero, the intercept gives $-(RT/F) \ln K_w$. The most reliable value of K_w obtained in this way is 1.008×10^{-14} at 25°C .

Ionic Product and Temperature.—The ionic product of water increases markedly with temperature, as is shown by the results in Table 165 (H.

TABLE 165. IONIC PRODUCT OF WATER

Temperature	0°	10°	25°	40°	50°	100° c.
K_w	0.113	0.292	1.008	2.917	5.474	$(56) \times 10^{-14}$

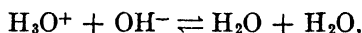
S. Harned and W. J. Hamer, 1933). Since the ionic product is the equilibrium constant of the reaction



it follows that application of the van't Hoff equation to the data in Table 165 will give the change in heat content accompanying the ionization of water. By expressing $\log K_w$ as a function of temperature and differentiating, the following values of ΔH have been obtained:

	0°	20°	60° c.
ΔH	14.51	13.69	11.94 kcal.

It has been seen that a strong acid in water can be considered as consisting of H_3O^+ ions, whereas a strong base consists of OH^- ions, and so the process of neutralization is actually



which is the reverse of the ionization process depicted above. The change in heat content for the neutralization of very dilute solutions of a strong acid and strong base should thus be equal, and opposite, to that for ionization of water. The best experimental value (T. W. Richards and A. W. Rowe, 1922), viz., -13.69 kcal. at 20° , for the heat of neutralization is in remarkable agreement with that calculated from the ionic product.

As a general rule, the heats of ionization * of weak acids and bases are relatively small, and so the dissociation constants do not change greatly with temperature; for example, K_a for acetic acid changes from 1.66×10^{-5} at 0° to 1.35×10^{-5} at 100° c. Since K_h is equal to K_w/K_a , or K_w/K_b , it is apparent that the hydrolysis constant must increase markedly as the temperature is raised; this accounts for the well known fact that if a salt solution is heated the extent of hydrolysis is increased.

It will be noted that whereas K_w increases with temperature, K_a for acetic acid appears to decrease. It is a matter of experience, however, that the dissociation constants of water, acids and bases increase to a

* The heat of ionization of a weak acid is given by the difference between the heat of neutralization of the acid, at an appreciable concentration when it is dissociated to a small extent only, and that of a strong acid in dilute solution, i.e., -13.69 kcal. at 20° c. A more accurate method (K. S. Pitzer, 1937) is to determine the heat evolved when a strong acid, assumed completely ionized, is added to a salt of the weak acid, also completely ionized, to yield the free acid, assumed to be completely nonionized. The heat of ionization of a weak base can be obtained in similar ways.

maximum and then decrease. For water the maximum is probably reached at about 250° c., but with acetic acid it is found to be at 25° c.

Ionic Product and Concentration.—The ionic activity product, $a_{H^+} \times a_{OH^-}$, should be constant in water or in a dilute solution at a given temperature, but the concentration product $c_{H^+} \times c_{OH^-}$, i.e., k_w , must vary with the ionic strength of the medium. As this increases, the product $f_{H^+} \times f_{OH^-}$ will at first decrease and then increase; hence k_w should increase, reach a maximum and then decrease with increasing amounts of electrolyte. Another factor which has been disregarded hitherto is the activity of the water: in calculating K_w it is assumed to be unity, but the assumption is justifiable only in dilute aqueous solutions. Since the presence of electrolytes lowers the vapor pressure, the activity of the water must decrease, and so the activity product K_w should increase. It is probably the compensation between the change in the activity coefficients of the hydrogen and hydroxyl ions and that of the water molecules which keeps the product k_w approximately constant in solutions containing appreciable amounts of electrolytes.³⁰

Determination of Hydrogen Ion Concentrations.—In principle it should be possible to determine the hydrogen ion activity or, approximately, concentration of a solution by measuring the potential of a hydrogen electrode inserted in the given solution. The E.M.F. of a cell, free from liquid junction potential, consisting of a hydrogen electrode and a reference electrode, should be given by

$$E = E_{\text{ref.}} - \frac{RT}{F} \ln a_{H^+}, \quad (230)$$

since the standard potential of the hydrogen electrode is zero. Introducing the definition of pH as $-\log a_{H^+}$, this becomes

$$E = E_{\text{ref.}} + 2.303 \frac{RT}{F} \text{pH};$$

$$\therefore \text{pH} = \frac{F(E - E_{\text{ref.}})}{2.303RT}. \quad (231)$$

By measuring the E.M.F. of the cell E obtained by combining the hydrogen electrode with a reference electrode of known potential, $E_{\text{ref.}}$, the pH of the solution may be evaluated. It has been found, however, that if the potentials of the reference electrodes given on page 941 are used, the results are not quite consistent with those obtained in other ways, e.g., from thermodynamic dissociation constants. This is due to uncertainties in connection with liquid junction potentials and single ion activities. Better agreement is obtained by taking the potential of the calomel reference electrode with 0.1 N potassium chloride solution, in conjunction with a saturated potassium chloride salt bridge, as 0.3358 volt at 25° c.

The hydrogen gas electrode is the ultimate standard of reference for

all pH determinations, but owing to the limitations and experimental difficulties associated with its use, other electrodes have been devised which are reversible with respect to hydrogen ions. The potentials of these electrodes are in agreement with (230) and (231), with the exception that $E_{ref.}$ must be changed in order to allow for the difference between the standard potentials of the hydrogen gas electrode and the other forms of hydrogen electrode. Various electrodes reversible with respect to hydrogen ions are described below.

Hydrogen Electrodes.—(i) *Hydrogen Gas Electrode.*—In the hydrogen gas electrode a small platinum sheet or wire is coated with platinum black (p. 891) and placed so that it is partly in the experimental solution and partly in an atmosphere of hydrogen at 1 atm. Hydrogen gas is bubbled through the solution; this is adsorbed by the platinized platinum which permits equilibrium between hydrogen molecules and the ions in solution to be established without any hindrance. The electrode thus behaves as if it were made of "metallic" hydrogen and gives a reversible potential depending on the activity of the hydrogen ions in solution. Since the standard potential is zero, the electrode potential gives a direct measure of the pH. For most purposes, especially for titrations, a convenient form of hydrogen electrode is that of J. H. Hildebrand (1913) shown in Fig. 241. A number of holes are made in the wider part of the tube at a level about midway up the platinum foil, so that when the electrode is inserted in a solution and hydrogen passed through the side-tube, the foil is half immersed in the liquid while the upper half is surrounded by hydrogen gas at 1 atm. pressure. Pure hydrogen should be used as the electrode is readily "poisoned." The hydrogen gas electrode is unsatisfactory in the presence of oxidizing agents, unsaturated organic compounds, alkaloids and the ions of noble metals.

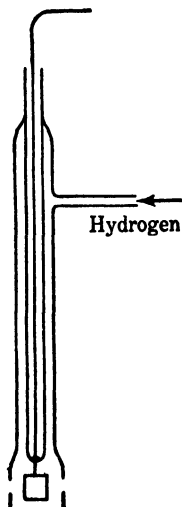


FIG. 241. Hydrogen electrode: Hildebrand type

(ii) *Oxygen and Oxide Electrodes.*—Theoretically the potential of an oxygen electrode, which is reversible with respect to hydroxyl ions, should be given by

$$E = E^0 + (RT/F) \ln a_{OH^-} = E^0 + (RT/F) \ln K_w - (RT/F) \ln a_{H^+},$$

and since K_w is a constant, this is of the same form as (230); the potential should thus be directly related to the pH of the solution. Many attempts have been made to set up an oxygen electrode similar to the hydrogen electrode described above, but the potential does not attain the theoretical value, and the changes with pH do not correspond to those required by (230). However, the potential does alter with the pH of the medium, and consequently oxygen and air electrodes, the latter being an oxygen electrode at 0.21 atm. pressure, have been used to find approximate end-points in the titration of solutions containing oxidizing agents, etc., which attack hydrogen. For similar purposes a suitable oxide electrode can be adopted; it consists of a metal and its oxide, e.g., $Sb|Sb_2O_3$, and it behaves like oxygen at a very small pressure,

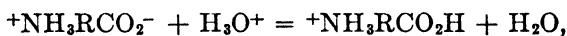
namely the dissociation pressure of the oxide. By casting antimony into stick form it is oxidized sufficiently to function as an oxide electrode; the potential responds reversibly to changes of hydrogen ion concentration according to (230) within the pH range of 2 to 8, and it gives approximate results even up to pH 11. The antimony electrode can, therefore, be used for potentiometric titrations of acids and bases (p. 1008). If it is required for the actual measurement of pH it must be standardized in solutions of known hydrogen ion concentration.

(iii) *The Quinhydrone Electrode.*—The potential of the quinone-hydroquinone system [equation (105)] depends on the ratio of the activities of the quinone and hydroquinone, and also on that of the hydrogen ions. If the compound quinhydrone, consisting of equimolecular amounts of Q and H_2Q , is dissolved in a solution then c_Q/c_{H_2Q} is unity, and hence a_Q/a_{H_2Q} is also approximately unity. In these circumstances (105) reduces to the general form of (230), so that the quinhydrone electrode behaves as a form of hydrogen electrode; according to experiments in solutions of known pH, the value of E° is $-0.6994 + 0.00074t$, where t is the temperature between 0° and 37° C. This may be combined with (231) to determine the pH of a given solution. The quinhydrone electrode is set up by adding a small amount of the compound to the solution and inserting a clean polished platinum or gold electrode. It behaves in a satisfactory manner under a variety of conditions, but it fails in solutions more alkaline than pH 8; this is partly due to the ready oxidation of the hydroquinone in alkaline solution and partly to its ionization (see p. 952). Provided the pH does not appreciably exceed 8, the quinhydrone electrode can be used for titration purposes.

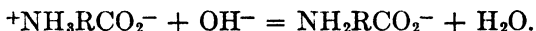
(iv) *Glass Electrode.*—It was discovered by F. Haber (1909) that the potential difference between a glass surface and a solution varied regularly with the pH of the latter; this observation has been made the basis of a method of measuring hydrogen ion concentrations. The glass electrode has attracted much attention in recent years because it can be used in almost any solutions, except those which are very acid or very alkaline. The electrode is made from a thin-walled bulb of low melting point glass of high electrical conductivity. In it is placed a solution of constant pH, e.g., a buffer solution (p. 1006), together with a little quinhydrone and a platinum wire to act as connection. The bulb is then inserted in the experimental solution and the difference of potential between the platinum wire and the external solution is determined. Since the resulting cell includes the glass bulb, the resistance is high, e.g., 10 to 100×10^6 ohms, and special methods (p. 926) have to be used to measure the potential. For actual pH determinations the glass electrode should be standardized in three or four solutions covering a range of hydrogen ion concentrations.²¹

Amphoteric Electrolytes.—The term "amphoteric" is applied to substances capable of exhibiting both acidic and basic functions; among these must, therefore, be included water, alcohols and other amphiprotic substances (p. 977), and a number of metallic hydroxides. In these compounds it is generally the same group which is responsible for the acidic and basic properties, but the **amphoteric electrolytes** or **ampholytes** containing *separate* acidic and basic groups are of special interest. The most familiar examples of this type of ampholyte are provided by the **amino-acids**, which may be represented by the formula NH_2RCO_2H ,

where R is aliphatic or aromatic in nature. For many years it was believed that the acidic properties of an amino-acid, as with carboxylic acids in general, was due to the $-\text{COOH}$ group, while the $-\text{NH}_2$ group exhibited the basic function. This is probably true for amino-benzoic acids and amino-phenols, but not for aliphatic amino-acids or aromatic amino-sulfonic acids. It was suggested by G. Bredig (1894) that a solution of an amino-acid might contain molecules carrying a positive charge at one end and a negative charge at the other, thus giving an electrically neutral system, viz., $^+\text{NH}_3\text{RCO}_2^-$. These particles have been variously called zwitterions, i.e., hybrid ions, amphions, ampholyte ions, dual ions or **dipolar ions**; the last has been used considerably in recent years. The concept of dipolar ions was employed by F. W. Küster (1897) to account for some of the properties of methyl orange, but the proportion of these ions present in solution was believed to be small. According to L.H. Adams (1916) and, particularly, N. Bjerrum (1923), however, nearly the whole of certain amino-acids, e.g., the aliphatic amino-acids, normally exists in solution in the form of the dipolar ions. An important consequence of this view is that the acidic function is due to the substituted ammonium ion, $^+\text{NH}_3-$, while the basic properties are those of the carboxylate ion, $-\text{CO}_2^-$. Thus, the reaction with a strong acid, represented by H_3O^+ , should be written as



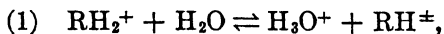
and that with a strong base, i.e., OH^- ion, should be



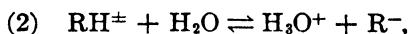
Evidence for Dipolar Ions.—The evidence for the presence of large proportions of dipolar ions in solutions of aliphatic amino-acids is very convincing. According to the older treatment the dissociation constants of the $-\text{NH}_2$ and $-\text{CO}_2\text{H}$ groups were extremely small, viz., about 10^{-8} to 10^{-12} ; such low values are difficult to understand, but they are not unreasonable when they refer to $-\text{CO}_2^-$ and $^+\text{NH}_3-$. It is known from studies on various acids and bases that in changing from water to alcohol as solvent the dissociation constants of carboxylic acids and amines are appreciably decreased, but those of cation acids, e.g., RNH_3^+ , and of anion bases, e.g., RCO_2^- , are hardly affected. Experiments with aliphatic amino-acids show that their strengths in alcohol are much the same as in water, but with aromatic amino-acids there are marked changes. It appears, therefore, that the former exist mainly as dipolar ions, whereas the latter ionize in the normal manner, the $-\text{NH}_2$ group exercising the basic function and the $-\text{COOH}$ group the acidic property. Formaldehyde is known to react with the $-\text{NH}_2$ group of an amino-acid; it is significant, therefore, that in the presence of formaldehyde the course of neutralization of an aliphatic ampholyte by an acid remains unchanged, whereas that by a base is altered. It is clear that the base is reacting with the nitrogen part of the molecule, as would be expected if the acidic group were the substituted ammonium ion, as postulated above. The basic neutralization of an aromatic amino-acid or of an amino-phenol is almost unaffected by formaldehyde, so that these substances behave like a normal combination of amine and carboxylic acid. Among the

properties of aliphatic amino-acids which favor the dipolar ion structure are high melting point, sparing solubility in nonpolar solvents, and high dielectric constant of aqueous solutions. Examination of glycine crystals by means of X-rays indicates the structure $^+\text{NH}_3\text{CH}_2\text{CO}_2^-$ in the solid state.

Dissociation Constants of Amino-Acids.—A very considerable simplification in the treatment of amino-acids can be achieved by regarding them as dibasic acids. Suppose the dipolar ion is represented in general by RH^\pm , e.g., $^+\text{NH}_3\text{CH}_2\text{CO}_2^-$, then the hydrochloride formed by the addition of an equivalent of acid is RH_2^+Cl^- , e.g., $(^+\text{NH}_3\text{CH}_2\text{CO}_2\text{H})\text{Cl}^-$. It is the positive ion RH_2^+ which behaves as a dibasic acid; the first stage of ionization is that of the $-\text{CO}_2\text{H}$ group, viz.,



and the second stage is due to the ammonium ion, viz.,



where R^- is $\text{NH}_2\text{CH}_2\text{CO}_2^-$ in the case of glycine. The dissociation constants for the two stages are then given by

$$K_1 = \frac{a_{\text{H}^+} \times a_{\text{RH}^\pm}}{a_{\text{RH}_2^+}} \quad \text{and} \quad K_2 = \frac{a_{\text{H}^+} \times a_{\text{R}^-}}{a_{\text{RH}^\pm}}, \quad (232)$$

the activity of the water being taken as unity. Of these two dissociation constants K_2 is the value for the acidic function of the amino-acid in dipolar form, i.e., for the $^+\text{NH}_3-$ group. On the other hand, K_1 is that for the acid *conjugate* to the basic function; hence, the dissociation constant for the latter, i.e., for the $-\text{CO}_2^-$ group, is equal to K_w/K_1 (cf. p. 978).

The dissociation constants K_1 and K_2 can be readily determined by E.M.F. methods which are quite analogous to those used for conventional dibasic acids; they are similar to those described for monobasic acids (p. 980). For glycine, for example, K_1 is 4.47×10^{-3} , so that the true basic dissociation constant of the dipolar form, in which it exists almost exclusively in pure aqueous solution, is $1.008 \times 10^{-14}/4.47 \times 10^{-3} = 2.24 \times 10^{-12}$ at 25° . The value of K_2 , giving the acidic dissociation constant, is 1.67×10^{-10} . These figures are reasonable for $-\text{CO}_2^-$ and $^+\text{NH}_3-$, respectively. Similar results have been obtained for other aliphatic amino-acids.

Hydrogen Ion Concentration in Ampholytes.—In any solution of a pure ampholyte there will be present to some extent the positive and negative ions, $^+\text{NH}_3\text{RCO}_2\text{H}$ and $\text{NH}_2\text{RCO}_2^-$, represented by RH_2^+ and R^- , respectively, in addition to hydrogen and hydroxyl ions and the neutral species $^+\text{NH}_3\text{RCO}_2^-$ and $\text{NH}_2\text{RCO}_2\text{H}$, i.e., RH^\pm and RH respectively. For most aliphatic acids the proportion of $\text{NH}_2\text{RCO}_2\text{H}$ is so small as to be negligible. Since the solution is always electrically neutral, the sum of the concentrations of the positive ions must equal

that of the negative ions, so that

$$c_{H^+} + c_{RH_2^+} = c_{R^-} + c_{OH^-}. \quad (233)$$

If the solution is sufficiently dilute for the activity coefficients to be taken as approximately equal to unity, the equations (232) may be written in the form

$$k_1 = c_{H^+} \times c_{RH^*}/c_{RH_2^+} \quad \text{and} \quad k_2 = c_{H^+} \times c_{R^-}/c_{RH^*}, \quad (234)$$

and if the values of $c_{RH_2^+}$ and c_{R^-} obtained from these relationships are inserted in (233) the resulting expression can be readily solved to give

$$c_{H^+} = \sqrt{\frac{k_1 k_2 c_{RH^*} + k_1 k_w}{k_1 + c_{RH^*}}}, \quad (235)$$

where k_w is introduced in order to convert c_{OH^-} into k_w/c_{H^+} . If the solution of the pure ampholyte is relatively acid, as is generally the case, the quantity c_{OH^-} in (233) can be neglected, and the value for c_{H^+} becomes

$$c_{H^+} = \sqrt{\frac{k_1 k_2 c}{k_1 + c}}, \quad (236)$$

where c_{RH^*} is replaced by c , the total concentration of the ampholyte, since it is almost exclusively in the dipolar ion form.

Isoelectric Points.—An ampholyte is at its **isoelectric point** when the concentration of the positive ions RH_2^+ is equal to that of the negative ions RH^- , i.e., $c_{RH_2^+} = c_{RH^-}$ at the isoelectric point. Since these ions are large they have almost the same equivalent conductance, and so equal amounts will migrate in opposite directions under the influence of an applied voltage. At the isoelectric point, therefore, an amino-acid, or a more complex ampholyte, will appear to remain stationary in an electrical field, although the solution may have an appreciable conductance. It should be made clear that in general it is not possible to obtain an isoelectric solution of a *pure* ampholyte; according to (233) if $c_{RH_2^+}$ is equal to c_{R^-} in a pure amino-acid, c_{H^+} must be equal to c_{OH^-} , and this would obviously be a rare occurrence. To attain the isoelectric point it is necessary, therefore, to add a small amount of acid or alkali; under these conditions (233) is no longer applicable. The hydrogen ion concentration of the isoelectric solution may be derived as follows. Neglecting activity coefficients, for simplicity, it is found from (234) that

$$c_{RH_2^+} = c_{H^+} \times c_{RH^*}/k_1 \quad \text{and} \quad c_{R^-} = k_2 c_{RH^*}/c_{H^+},$$

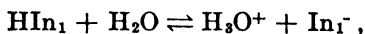
and if these are equated, the condition for the isoelectric point is seen to be

$$c_{H^+} = \sqrt{k_1 k_2}. \quad (237)$$

It can be shown that at this hydrogen ion concentration the total extent of ionization of the amino-acid as acid and base is a minimum. For this

reason many of the physical properties of ampholytes, such as solubility and viscosity, exhibit maximum or minimum values at their respective isoelectric points.³²

Neutralization Indicators.—A neutralization indicator is a substance which, within certain limits, varies in color according to the hydrogen ion concentration of its environment; it invariably possesses, in some form or another, the properties of a weak acid or weak base. As a result of investigations in organic chemistry, it was suggested that an indicator consists of two or more tautomeric forms having different constitutional formulae and different colors. At least one and frequently more of these tautomers are acidic or basic, and the color of each ion is the same as that of the nonionized form of the same structure.* Suppose HIn_1 is the nonionized colorless form of the indicator; it will ionize as an acid to give the colorless In_1^- ion, thus



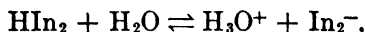
and the equilibrium constant, using concentrations, is

$$k_1 = c_{\text{H}^+} \times c_{\text{In}_1^-} / c_{\text{HIn}_1} . \quad (238)$$

The colorless ion In_1^- will be in tautomeric equilibrium with the colored ion In_2^- having a different structure, viz., $\text{In}_1^- \rightleftharpoons \text{In}_2^-$, so that the corresponding equilibrium constant is

$$k_T = c_{\text{In}_2^-} / c_{\text{In}_1^-} . \quad (239)$$

Finally, the colored ions In_2^- will be in equilibrium with hydrogen ions and the colored nonionized molecules HIn_2 , i.e.,



so that

$$k_2 = c_{\text{H}^+} \times c_{\text{In}_2^-} / c_{\text{HIn}_2} . \quad (240)$$

Combination of (238), (239) and (240) gives

$$\frac{c_{\text{H}^+}(c_{\text{In}_1^-} + c_{\text{In}_2^-})}{(c_{\text{HIn}_1} + c_{\text{HIn}_2})} = \frac{k_1 k_2 (1 + k_T)}{k_2 + k_1 k_T} = k_{\text{In}} , \quad (241)$$

where k_{In} is a composite constant; it follows, therefore, that

$$c_{\text{H}^+} = k_{\text{In}} \frac{(c_{\text{HIn}_1} + c_{\text{HIn}_2})}{(c_{\text{In}_1^-} + c_{\text{In}_2^-})} . \quad (242)$$

If a substance is to be a satisfactory indicator for changes of hydrogen ion concentration, the numerator and denominator of (242) must correspond to two distinct colors; this can occur only if the nonionized molecules are almost completely either HIn_1 or HIn_2 , and the ions must be almost exclusively in the other form. In other words, k_T must be very

* In view of the modern theory of color as being associated with resonance (p. 584), these suggestions require modification; the results obtained, however, remain unaffected.

small and k_1/k_2 large, or *vice versa*. If this were not the case, both numerator and denominator of (242) would represent a somewhat similar mixture of colors, and although a change in c_{H^+} would alter the ratio of numerator to denominator, the color might be almost unchanged. For a satisfactory indicator (242) may be reduced to

$$c_{H^+} = k_{In} \frac{[\text{Nonionized form}]}{[\text{Ionized form}]} = k_{In} \frac{1 - \alpha}{\alpha}, \quad (243)$$

where α is the fraction of the total indicator present as ions. The actual color, which depends on the ratio of the two forms, is thus directly related to the hydrogen ion concentration. In an acid medium, i.e., c_{H^+} is high, the concentration of nonionized form must increase, whereas in an alkaline environment the ionized form will be produced.

Indicator Range.—It may be assumed that when less than about 9 per cent of the total indicator is ionized, i.e., $\alpha = 0.09$, the color of the ionized form is just barely visible. The lowest hydrogen ion concentration at which the indicator still shows its acid color, e.g., due to HIn_1 , will thus be given by

$$c_{H^+} = k_{In} \frac{0.91}{0.09} \approx 10k_{In}; \quad \therefore pH = pk_{In} - 1. \quad (244)$$

On the other hand, when 91 per cent of the indicator is in the ionized form, $\alpha = 0.91$, i.e., the color corresponds to the almost completely alkaline condition due to In_2^- . The highest hydrogen ion concentration at which the indicator shows its full alkaline color is then

$$c_{H^+} = k_{In} \frac{0.09}{0.91} \approx \frac{1}{10}k_{In}; \quad \therefore pH = pk_{In} + 1. \quad (245)$$

Hence, as the pH of a solution is increased by adding alkali the color of an indicator begins to change at a pH equal to $pk_{In} - 1$, and is completely changed, as far as the eye can detect, at a pH of $pk_{In} + 1$. The transition interval of an indicator is thus *very approximately* one unit of pH on each side of its particular pk_{In} . When c_{H^+} is equal to k_{In} , i.e., $pH = pk_{In}$, the indicator color is exactly intermediate between its acid and alkaline forms.

Although the above discussion has had particular reference to acid indicators, it can be applied to basic indicators. As a matter of convenience any base can be regarded as equivalent to its conjugate acid, and treated as an acid of k_a equal to k_w/k_b . All indicator systems, of course, consist of a conjugate acid and base, viz., HIn and In^- , and it is in a sense somewhat arbitrary to refer to certain indicators as acids and to others as bases. The particular term employed is determined by the nature of the form in which it is usually encountered, and has no real significance as far as indicator properties are concerned.

Indicator constants are determined by placing the substance in a solution of known hydrogen ion concentration and finding, by a colorim-

eter or spectrophotometer, the ratio of the amounts of the two forms having different colors; k_{In} can then be evaluated by (243). Alternatively, the hydrogen ion concentration may be determined at which the indicator shows its intermediate color: two solutions, one acid and the other alkaline, are superimposed, and a solution found in which the indicator shows the same color. The c_{H^+} in this solution is then equal to k_{In} . In some cases, e.g., *p*-nitrophenol and methyl orange, the value of the indicator constant, which is almost identical with the ordinary dissociation constant, may be derived from conductance or hydrolysis measurements. The properties of a number of useful indicators are given in Table 166

TABLE 166. USEFUL INDICATORS AND THEIR CHARACTERISTICS

Indicator	Color		pK_{In}	pH Range
	Acid	Alkaline		
Thymol blue	Red	Yellow	1.51	1.2-2.8
Bromophenol blue	Yellow	Blue	3.98	3.0-4.6
Chlorophenol red	Yellow	Red	5.98	4.8-6.4
Bromothymol blue	Yellow	Blue	7.0	6.0-7.6
Cresol red	Yellow	Red	8.3	7.2-8.8
Thymol blue (2nd range)	Yellow	Purple	8.9	8.0-9.6
Methyl orange	Red	Yellow	3.7	3.1-4.4
Methyl red	Red	Yellow	5.1	4.2-6.3
Phenolphthalein	Colorless	Red	9.4	8.3-10.0

(W. M. Clark, 1928). The ranges recorded are those found by experiment; they show that the rule derived previously is only approximate, for the range of visible color change is not necessarily 9 to 91 per cent.

Determination of pH by Indicators.—The best method for determining hydrogen ion concentrations with indicators requires the availability of a number of solutions of known hydrogen ion concentration (p. 1007). After a preliminary experiment, for example with a mixed indicator, sometimes called a "universal indicator," to obtain an approximate value of the pH of the given solution, the appropriate indicator is chosen. A definite amount is then placed in a certain volume of the test solution, and the color compared with those produced, at the same concentration, in a series of solutions of known pH. By matching the colors the pH of the unknown solution may be determined within 0.05 to 0.1 unit. If solutions of known hydrogen ion concentration are not available, then application may be made of (243) written in the form

$$c_{H^+} = k_{In} \frac{\text{Color due to acid form}}{\text{Color due to alkaline form}} \quad (246)$$

or

$$pH = pK_{In} - \log \frac{\text{Color due to acid form}}{\text{Color due to alkaline form}} \quad (247)$$

Provided the indicator constant is known, it is only necessary to determine the ratio of acid to alkaline color of the indicator in the given solution for its pH to be calculated by (247). The ratio of the colors is obtained most accurately

by a colorimeter or by means of a spectrophotometer, but for approximate purposes various simple devices have been employed. In N. Bjerrum's "wedge method" (1914), for example, a rectangular glass box is divided by a diagonal glass sheet into two wedge-shaped compartments; one is filled with acidified indicator and the other contains the alkaline form. By viewing the combination from the front a series of colors are observed representing a continuously varying ratio of acid and alkaline forms. The color of the test solution with the indicator can then be matched and the ratio of the depths of the two sections of the wedge gives the information whereby the pH may be evaluated from (247). In the "drop-ratio" method of L. J. Gillespie (1920) the ratio of acid to alkaline colors in the solution being examined is determined by comparison of its color with those of a series of pairs of tubes, one of each pair containing a definite number of drops of alkaline indicator and the other of acid form of the indicator.

If reasonably accurate results are required it is necessary to take into account the effect of temperature on the indicator constant, and particularly the **neutral salt error**. A change in the ionic strength of the medium may produce an appreciable change in the "classical" dissociation constant, but the necessary corrections can be made by equations of the same form as (166) or (166a). When working with protein solutions care should be taken to ensure that there is no combination between the protein and the indicator.

Neutralization Curves.—The changes in hydrogen ion concentration accompanying the addition of a base to an acid are important for analytical purposes and the subject will be considered here mainly from the standpoint of aqueous solutions. A dilute solution of a strong acid may be assumed to be completely ionized, and the concentration of hydrogen ions is equal to the concentration of un-neutralized acid at any instant. If a is the initial concentration of acid in g. equiv. per liter and y is the fraction neutralized then c_{H^+} is equal to $a(1 - y)$ g. ion per liter. At the equivalence point, that is, when equivalent amounts of acid and alkali have been mixed, the solution, being that of a salt of a strong acid and strong base, is neutral and so its pH is 7. The variation of hydroxyl ion concentration, and hence of hydrogen ions, since $c_{H^+} \times c_{OH^-} = 10^{-14}$, during the neutralization of a strong base by a strong acid can be determined in a similar manner. The complete results for 0.1 N solutions of strong acid and strong base are plotted as the curve IA-IB in Fig. 242 (p. 1004); it shows the variation of the pH during the course of neutralization. The attainment of the equivalence point, i.e., the theoretical end-point of the titration, is seen to be accompanied by a sudden change of pH.

For a weak acid and strong base the calculation is more complicated because (i) the acid is only slightly ionized, (ii) the salt formed during neutralization represses the ionization of the acid, and (iii) the salt solution is not neutral because of hydrolysis. Let a represent the total initial concentration of weak acid HA ; during neutralization by a completely ionized strong base, e.g., $NaOH$, it is converted into the salt NaA , also completely ionized. If b is the concentration of base added at any instant, then c_{Na^+} is equal to b , whereas a is equal to $c_{HA} + c_{A^-}$. Since the solution

is electrically neutral

$$c_{Na^+} + c_{H^+} = c_{A^-} + c_{OH^-}, \quad (248)$$

and from this and the values of a and b , it is possible to eliminate c_{A^-} and c_{HA} from the expression for the classical dissociation constant, i.e.. $k_a = c_{H^+} \times c_{A^-}/c_{HA}$, giving

$$c_{H^+} = k_a \frac{a - b - c_{H^+} + c_{OH^-}}{b + c_{H^+} - c_{OH^-}}. \quad (249)$$

Apart from activity corrections, which should strictly be introduced, (249) gives c_{H^+} from the beginning to the end of neutralization; it is, however, a cubic equation, since $c_{OH^-} = k_w/c_{H^+}$, and so cannot be solved easily. Simplifications can, however, be made; for example, as long as the solution is acid, c_{OH^-} can be neglected, and so

$$c_{H^+} = k_a \frac{a - b - c_{H^+}}{b + c_{H^+}}. \quad (250)$$

This is a quadratic from which c_{H^+} can be readily calculated for various amounts of acid and base. Since a is the original concentration of acid and b that of the base added, then $a - b$ is the concentration of acid remaining unneutralized, and b is the amount of salt formed; hence (250) may be written

$$c_{H^+} = k_a \frac{[\text{acid}] - c_{H^+}}{[\text{salt}] + c_{H^+}} \quad (251)$$

or

$$pH = pk_a + \log ([\text{salt}] + c_{H^+})/([\text{acid}] - c_{H^+}), \quad (252)$$

where the square brackets represent concentrations. If the solution is not too acid, then c_{H^+} may be neglected, and (252) becomes

$$pH = pk_a + \log [\text{salt}]/[\text{acid}]. \quad (253)$$

This is the equation of L. J. Henderson (1908), which is usually known by his name. It gives very satisfactory results for solutions between pH 4 and pH 10, provided they are not too dilute; for pH values less than 4, equation (252) must be employed.

The Henderson equation may be used to calculate the pH during the neutralization of most acids, but it is not applicable at the beginning and end; these values may, however, be readily obtained in another way, so that almost the whole of the pH -neutralization curve can be easily mapped. Before the commencement of neutralization b is zero, and c_{OH^-} can be neglected, and so it follows from (249) that

$$c_{H^+} = k_a(a - c_{H^+})/c_{H^+}; \quad (254)$$

$$\therefore c_{H^+} = -\frac{1}{2}k_a + \sqrt{\frac{1}{4}k_a^2 + ak_a}. \quad (255)$$

The same result may be obtained by considering a solution containing acid only. The concentration of nonionized molecules, c_{HA} , is equal to the total concentration a less the amount ionized, c_{H^+} ; hence $c_{HA} = a - c_{H^+}$.

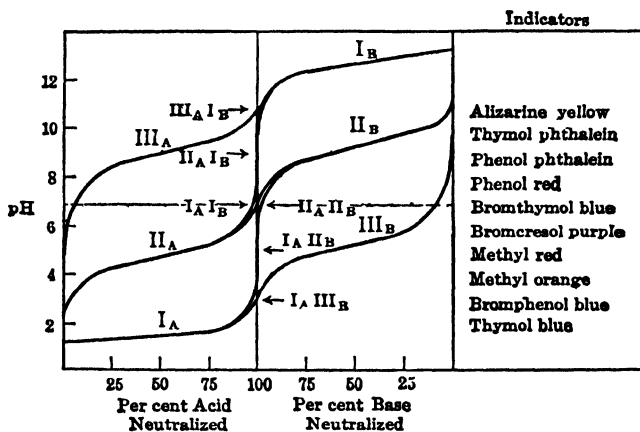


FIG. 242. Neutralization curves

Further, the concentration of H^+ and A^- ions must be equal, and so $c_{A^-} = c_{H^+}$; substitution in the classical equation $k_a = c_{H^+} \times c_{A^-} / c_{HA}$ gives (254). If k_a is small, i.e., for a fairly weak acid, (255) reduces to the useful form

$$c_{H^+} \approx \sqrt{ak_a}. \quad (256)$$

At the end of neutralization or, more exactly, at the equivalence point, a and b are equal, so that $a - b$ is zero. Since the solution is alkaline as a result of hydrolysis c_{H^+} in the numerator of (249) may be neglected, and so also may $c_{H^+} - c_{OH^-}$, in comparison with b in the denominator. Hence (249) gives for the end-point

$$c_{H^+} = k_a c_{OH^-} / a = \sqrt{k_w k_a / a}, \quad (257)$$

where c_{OH^-} is replaced by k_w / c_{H^+} . It will be noted that (257) is identical, as it should be, with (213), for the hydrogen ion concentration of the solution of a salt of a weak acid and strong base.

Similar equations to those given above may be derived for the neutralization of a weak base by a strong acid; it is merely necessary in each case to interchange a and b , k_a and k_b , and c_{H^+} and c_{OH^-} . The corresponding form of the Henderson equation, applicable at pH values less than 10, is

$$\begin{aligned} pOH &= pk_b + \log [\text{salt}] / [\text{base}]; \\ \therefore pH &= pk_w - pk_b - \log [\text{salt}] / [\text{base}]. \end{aligned} \quad (258)$$

Strictly speaking, special equations should be used for the neutralization

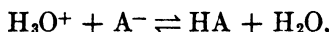
of a weak acid and weak base, but the Henderson equation is applicable over nearly the whole of the range. It is only near the equivalence point that the results diverge, but since the pH at this point can be derived from (222) the discrepancy is not serious.

By means of these various equations the pH -neutralization curves for acids and bases of different types can be calculated. The curve IIA-IB in Fig. 242 represents the neutralization of a 0.1 N solution of a weak acid with k_a equal to 1.0×10^{-5} by a strong base; IIIA-IB is for a very weak acid, e.g., a phenol, of $k_a = 10^{-9}$, and a strong base; IA-IIIB and IA-IIIIB are the pH curves for a strong acid and a weak base ($k_b = 10^{-5}$) and a very weak base ($k_b = 10^{-9}$), respectively; finally, IIA-IIIB is for a weak acid and weak base, both of whose dissociation constants are 10^{-5} . The position of the equivalence point is indicated by an arrow in each case.

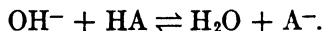
Indicators in Neutralization Titrations.—A study of pH -neutralization curves is of great importance in analytical chemistry, because it permits of a correct choice of indicator for any particular titration. It is necessary, of course, to use an indicator which changes color in the region of pH in the vicinity of that at the equivalence point. In the titration of 0.1 N strong acid and 0.1 N strong base, for example, the pH of the solution undergoes a very sharp change from pH 4 to 10 within 0.1 per cent of the equivalence point (Fig. 242, IA-IB). Any indicator changing color in this range can, therefore, be used to give a reliable indication when the end-point is reached. Consequently, both phenolphthalein, which just turns pink at pH 9.0, and methyl orange, changing visibly at pH 4.0, may be employed. If the solutions are diluted to 0.01 N, however, the change in pH at the equivalence point is less sharp, viz., from 5 to 9; methyl orange will then give results which are too low, for the color will change at pH 4, that is, before the acid is neutralized. When 0.1 N acetic acid is titrated with 0.1 N alkali the pH at the equivalence point is 8.87, and there is a fairly sharp increase from pH 7.8 to 10 (Fig. 242, IIA-IB); of the common indicators phenolphthalein only is satisfactory, or alternatively the less familiar thymol blue (second range) may be used. Any indicator changing below about pH 8 will be unsatisfactory for two reasons: first, the color will change before the equivalence point, and second, the change will not be sharp because of the relatively gradual change of pH below this value. For the titration of 0.1 N ammonia with a strong acid, the indicator transition should occur at about pH 5.1, and the best is methyl red, changing at pH 5; although methyl orange is often used, the results are in error, unless the solutions are at least 0.1 N. It should be evident that for an indicator color to change sharply at the end-point the pH -neutralization curve must rise rapidly at this point, otherwise there will be a gradual drift from alkaline to acid color, or *vice versa*. No satisfactory indicator can therefore be found for the titration of a very weak acid (Fig. 242, IIIA-IB), of a very weak base (IA-IIIIB), or of a moderately weak acid by a similar base (IIA-IIIB). Approximate results can some-

times be obtained by using a suitable indicator and titrating to the theoretical pH by means of a comparison flask containing a solution, together with indicator, at that pH .²³

Buffer Solutions.—Aqueous solutions of both sodium chloride and of ammonium acetate have a pH of about 7, but the addition of 1 cc. of 0.1 N hydrochloric acid to 1 liter alters the pH to 4 in the former case, although it hardly affects the latter. The resistance to change in the hydrogen ion concentration of a solution on the addition of acid or alkali is known as **buffer action**, and a **buffer solution** almost invariably consists of a mixture of a weak acid, or a weak base, and its salt. Buffer action in a solution of a weak acid (HA) and its salt (A^-) is explained by the fact that hydrogen ions are "neutralized" by the anions of the salt, acting as a weak base; thus,



whereas hydroxyl ions are removed by the neutralization



If the buffer is a mixture of a weak base (B) and its salt (BH^+), the corresponding equations are



The buffer capacity of a solution is defined by db/dpH , where dpH is the increase of pH resulting from the addition of db of base; the addition of acid, which decreases the pH , is equivalent to a negative amount of base. The greater the buffer capacity the larger the amount of acid or alkali which must be added to produce a definite change of pH . An examination of Fig. 242 shows that the slope of the pH -neutralization curves is in fact the reciprocal of the buffer capacity, and so the curves are flattest in the regions of high buffer capacity, but where they rise rapidly the buffering is poor. It is seen from curves Ia and Ib that a relatively concentrated solution of a strong acid or strong base can act as a buffer in regions of low or high pH , respectively; this is because the salt does not repress the ionization of the acid or base. A solution of a weak acid or base *alone* is a poor buffer, as may be seen from the marked change of pH at the commencement of neutralization in each case. When an appreciable amount of salt is present, that is to say, towards the middle of the curves IIa or IIIa, for a weak acid and its salt, or IIb or IIIb for a weak base and its salt, the buffer capacity is marked. It can be shown that it is actually greatest, as may be anticipated, at the middle of the curve, when the solution consists of equivalent amounts of the acid, or base, and its salt. At this point the pH is equal to pK_a of the acid, or pOH to pK_b of the base (p. 980), and the appreciable buffer properties of an acid (or base) and its salt generally extend over a pH (or pOH) range equal to $pk + 1$ to $pk - 1$. The buffer capacity of a salt is generally small, as is evident from the sharp increase of pH near the equivalence point; a salt

of a weak acid and a weak base, or that of an extremely weak acid or base, is, however, somewhat exceptional. The neutralization curves are seen to be relatively flat at the equivalence point so that the solutions, e.g., ammonium acetate, have appreciable buffer capacity; this is due to hydrolysis and the presence of considerable amounts of free acid and free base in a solution of the salt.

For many aspects of chemical work, e.g., standardization of indicators, hydrogen electrodes, etc., it is desirable to have solutions of known hydrogen ion concentration which are not sensitive to the influence of acidic or basic impurities, and buffer solutions are employed for the purpose. To make a buffer solution of a given pH , it is first necessary to choose an acid (or base) with a pK value as near as possible to the required pH (or pOH) for the buffer capacity to be large in this region. The actual ratio of acid to salt necessary can then be found from the Henderson equation (253) or (258), e.g.,

$$pH = pK_a + \log [\text{salt}]/[\text{acid}].$$

As derived on p. 1003 this equation applies to a partly neutralized acid, but such a system is obviously equivalent to a mixture of the acid and its salt. If the required pH is less than 4 or greater than 10 the more accurate equations should be employed. Sometimes a buffer solution is made up of two salts representing different stages of neutralization of a polybasic acid, e.g., NaH_2PO_4 and Na_2HPO_4 ; in this case the former is the acid and the latter is the salt, the dissociation constant being that of the acid $H_2PO_4^-$.

For accurate work it is desirable that the pH of a buffer solution should be checked by means of the hydrogen electrode, because in the equations derived above activity coefficients have been neglected and the dissociation constant is assumed to be the classical value. Provided the ionic strength is not too great the necessary corrections could be made by the limiting Debye-Hückel equation. In view of the importance of buffer mixtures a number of such solutions have been made up and their pH values carefully checked experimentally. By following the directions given a solution of any desired pH can be prepared with rapidity and precision. A few of the mixtures studied, and their effective ranges, are given in Table 167. Each buffer system is generally applicable over a

TABLE 167. BUFFER MIXTURES

Mixture	Range	Mixture	Range
Phthalic acid and Potassium acid phthalate	2.2-3.8	Boric acid and Borax	6.8-9.2
Phenylacetic acid and Sodium phenylacetate		Diethylbarbituric acid and Sodium salt	
Potassium acid phthalate and Dipotassium phthalate	3.2-4.9	Borax and Sodium hydroxide	7.0-9.2
Sodium dihydrogen phosphate and Disodium hydrogen phosphate		Disodium hydrogen phosphate and Sodium hydroxide	
	4.0-6.2		9.2-11.0
	5.9-8.0		11.0-12.0

limited range, but by making suitable mixtures of acids and acid salts it is possible to prepare "universal buffer mixtures"; by adding various predetermined amounts of alkali, buffer solutions of any desired pH from 2 to 12 may be obtained. An example of this type is a mixture of citric acid, diethylbarbituric acid, boric acid and potassium dihydrogen phosphate; this is virtually a system of seven acids, and so, when suitably neutralized, is capable of exercising appreciable buffer capacity in seven pH regions.³⁴

POTENTIOMETRIC TITRATIONS

Neutralization Titrations.—Instead of employing an indicator, the end-point in a neutralization titration may be found by direct measurement of the pH of the solution after each addition of alkali to acid; that is to say, the pH -neutralization curve is actually followed experimentally and hence the equivalence point can be found. It can be shown theoretically that, as a general rule, the change of pH for the addition of a given amount of titrant is a maximum at the equivalence point, and so the latter can be identified. The determination of pH is carried out with some form of hydrogen electrode, the potential of which is measured; hence the process is called **potentiometric titration**. The potential of any hydrogen electrode (p. 994), is given by

$$E = E^0 - 0.0002T \log a_{H^+} = E^0 - 0.059 \log a_{H^+} \quad (259)$$

at room temperature, where E^0 is a constant for the particular electrode. Since pH is defined as $-\log a_{H^+}$, it follows that $E = E^0 + 0.059 pH$; the change in electrode potential, or of the E.M.F. of a cell made up of the hydrogen and a standard electrode, is thus proportional to the change in pH during titration. At the end-point, therefore, $\Delta E/\Delta V$ is a maximum, where ΔE is the change in potential or E.M.F. resulting from the addition of a definite volume, e.g., 0.1 cc. of titrant. The accuracy with which this point can be detected depends on the magnitude of the inflexion in the pH -neutralization curve. As seen in Fig. 242, this is more marked the stronger the acid and base; nevertheless it is possible to determine potentiometrically an approximate end-point in the titration of very weak acids or very weak bases when indicator methods are quite unreliable. The potentiometric method can also be employed for the analysis of dilute or highly colored solutions.

Oxidation-Reduction Titrations.—The completion of an oxidation-reduction titration, e.g., of a ferrous salt by acid dichromate or permanganate, is accompanied by an increase of potential similar to that at the end-point of a neutralization reaction. Provided the standard potentials of the two systems involved, i.e., the one oxidized and the one reduced, differ by at least 0.3 volt, the change of potential at the end-point is sharp and can be readily detected. The experimental method is very simple, since it is necessary only to insert a clean platinum electrode into the

solution and to measure its potential in the course of the titration. It can be shown theoretically that the maximum of $\Delta E/\Delta V$ does not occur exactly at the equivalence point, but the discrepancy is smaller the more concentrated the solutions and the greater the difference in the standard potentials of the reacting systems.

Until the recent development of oxidation-reduction indicators for titration purposes the potentiometric method was the only one available for the determination of the end-points in certain volumetric processes. These indicators, like those mentioned on p. 953, have different colors in oxidized and reduced states, and they adjust themselves to the same potential as the medium in which they are placed. Since the potential of the latter changes rapidly at the equivalence point of a titration, the color of the indicator will also change provided it is suitably chosen. The standard potential of the indicator must obviously lie between the values for the two systems being titrated if it is to show a sharp color transition.

Precipitation Titrations.—When a solution of one ion, e.g., silver, is titrated with that of another, e.g., chloride, with which it forms a sparingly soluble compound, the concentration of the former ion decreases during the process, and this decrease is relatively most rapid near the equivalence point. The potential of a silver electrode placed in the titrated solution will change in an analogous manner in the course of the titration. If the two solutions are of the same concentration and the precipitate is formed from ions of the same valence, e.g., as in silver chloride, the maximum value of $\Delta E/\Delta V$ occurs exactly at the equivalence point. Should the concentrations or valences be different, however, then this is not strictly true, but the deviations are small provided the solutions are not too dilute and the precipitate is not very soluble. These are also the conditions for an appreciable inflexion in the potential-titration curve, and so when the inflexion point can be detected easily it must also be close to the theoretical end-point. The potentiometric method is particularly useful for the titration of halides, cyanides, and other anions forming sparingly soluble silver salts; it can be adapted to the study of mixtures of halides and of halides with cyanide.

Various methods have been devised for the simplification of potentiometric titrations, but for an account of them the specialized literature must be consulted.¹¹

THERMODYNAMICS OF ELECTROLYTES AND IONS

Thermal Properties.—The electrical energy of a concentration cell without transference, is equal to the free energy decrease ($-\Delta F$) accompanying the transfer of 1 mole of electrolyte from a more concentrated solution, molality m , to the dilute solution m' (p. 927). This may be represented by $zF(E' - E)$ where E and E' are the E.M.F.'s of the two half-cells and z is the valence of the ions with respect to which the extreme electrodes are reversible. The increase of heat content (ΔH) accompanying the transfer process is $\bar{H}_2 - \bar{H}'_2$, where \bar{H}_2 and \bar{H}'_2 are the partial molar heat contents of the solute in the two solutions.

Insertion of these values of ΔH and ΔF in the Gibbs-Helmholtz equation (46) gives

$$\bar{H}_2 - \bar{H}'_2 = zF \left\{ (E' - E) + T \frac{\partial(E' - E)}{\partial T} \right\}. \quad (261)$$

If the E.M.F. of the concentration cell is measured at various temperatures, at constant pressure, it is possible to determine the change in heat content $\bar{H}_2 - \bar{H}'_2$. It has been seen (p. 242) that the quantity $\bar{H}_2 - \bar{H}_2^0$, where \bar{H}_2^0 refers to an infinitely dilute solution, is equal to \bar{L}_2 , the relative partial molar heat content, which is a quantity of thermodynamic interest. To evaluate this directly from (261) it would be necessary for one of the half-cells to contain the electrolyte at infinite dilution; this is, however, not feasible and so extrapolation methods are employed.

Combination of (148) and (150a) gives for the chemical potential of an electrolyte

$$\mu_2 = \mu_2^0 + \nu RT \ln a_{\pm} \quad (262)$$

$$= \mu_2^0 + \nu RT \ln m + \nu RT \ln \gamma, \quad (263)$$

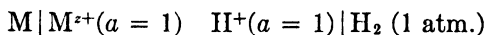
where ν is the number of ions produced by one molecule of electrolyte, and m and γ are the *mean* molality and activity coefficient, respectively, of the ions. By making use of the Gibbs-Helmholtz equation in the form of (96), p. 686, it can be readily deduced that

$$\bar{L}_2 = \bar{H}_2 - \bar{H}_2^0 = -\nu RT^2 \frac{\partial \ln \gamma}{\partial T} \quad (264)$$

at constant pressure. The value of \bar{L}_2 may thus be derived from the variation of the mean activity coefficient of an electrolyte with temperature. According to (189), p. 243, the relative partial molar heat capacity is equal to the temperature coefficient of \bar{L}_2 and so the former can be evaluated by a second differentiation of the activity coefficient with respect to temperature.

The limiting Debye-Huckel equation requires the activity coefficient to be a linear function of \sqrt{c} in dilute solution; it follows, therefore, that both the relative partial molar heat content and heat capacity should vary in a linear manner with \sqrt{c} . The experimental data from E.M.F. measurements bear out this expectation, but the solutions used are probably too concentrated for the limiting law to be applicable and so the slopes of the lines differ from the theoretical values.

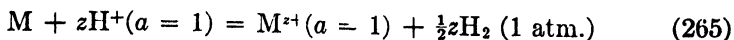
Standard Free Energy of Ions.—The chemical potential of any ionic species is $\mu = \mu^0 + RT' \ln a$, and if the ions are at unit activity, that is, in their standard states, μ is equal to μ^0 . In the hypothetical cell



with the ions M^{z+} and H^+ at unit activity, the potential of the left-hand electrode is E^0 , the standard potential,* and that of the right-hand electrode is zero by convention. The decrease of free energy in the cell is thus zE^0F , where z is the valence of the M^{z+} ions. The chemical reaction in

* Unless otherwise stated, the standard potential always implies the standard oxidation potential.

the cell is



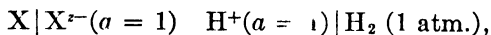
and the decrease of free energy, $-\Delta F^0$, is given by

$$zE^0F = -\Delta F^0 = \mu_M^0 + z\mu_{H^+}^0 - \mu_{M^{z+}}^0 - \frac{1}{2}z\mu_{H_2}^0, \quad (266)$$

all the substances being in their standard states. If the free energy change accompanying the formation of hydrogen ions at unit activity from the gas at 1 atm. pressure, i.e., $\mu_{H^+}^0 - \frac{1}{2}\mu_{H_2}^0$, is arbitrarily taken as zero, then

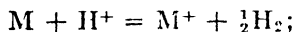
$$zE^0F = -\Delta F^0 = \mu_M^0 - \mu_{M^{z+}}^0, \quad (267)$$

which is the decrease of free energy accompanying the formation from M of M^{z+} ions at unit activity. The corresponding value of ΔF^0 , which is equal to $-zE^0F$, is called the **standard free energy** of formation of the particular ions, although it is really the increase of free energy of the reaction (265). Similarly, by considering a cell involving negative ions, viz.,



and making the same assumption that the free energy of formation of hydrogen ions is zero, it can be shown that the standard free energy of formation of an anion is $+zE^0F$. The free energies obtained by expressing E^0 in volts and F , as usual, in coulombs, will be in joules and multiplication by 0.2390 is necessary for conversion into calories.

Standard Entropies of Ions.—Since 1923, W. M. Latimer and his collaborators have made a valuable collection of the standard partial entropies of ions in aqueous solution. The quantity called the **standard entropy** is really the relative partial entropy of the ion in a solution of unit activity, that of the hydrogen ion being assumed zero; it is given the symbol \bar{S}_i^0 , the nature of the ion being indicated by the subscript. The simplest method of evaluating the standard entropy is the following. Consider the reaction



the free energy change ΔF^0 is equal to $-E^0F$, as seen above, and ΔH^0 is obtained from the heat of solution of 1 g. atom of the metal M in a large excess of a dilute solution of hydrogen ions, so that the metal (M^+) ions produced are also at high dilution. Applying the familiar thermodynamic equation $\Delta S^0 = (\Delta H^0 - \Delta F^0)/T$ it is possible to evaluate ΔS^0 for the reaction; this is equal to

$$\Delta S^0 = \bar{S}_{M^+}^0 + \frac{1}{2}\bar{S}_{H_2} - \bar{S}_M - \bar{S}_{H^+}^0. \quad (268)$$

The entropies of hydrogen gas and of the metal M are known (see Chapter XI), $\bar{S}_{H^+}^0$ is assumed to be zero, and so $\bar{S}_{M^+}^0$ can be evaluated. For example, when the metal M is potassium, ΔF^0 , from the standard potential, is $-67,630$ cal. and ΔH^0 from thermochemical observations is $-60,270$

at 25° c., and so ΔS is 24.6 E.U. The entropy of $\frac{1}{2}\text{H}_2$ is 15.6 (Table 148) and that of K is 15.2, so that $\bar{S}_{\text{K}^+}^0$ is 24.2 E.U. per g. ion at 25°. If the metal is noble and does not dissolve in acid ΔH can be calculated from two thermal measurements: one of the heat of solution of a base metal and the other of the heat of displacement of the noble by the base metal.

For anions, e.g., NO_3^- or SO_4^{--} , whose standard potentials are not known, the following method is employed. In a saturated solution of a salt the chemical potential of the salt in the solution, as given by (262), must be equal to that of the solid salt; that is

$$\mu_s = \mu_2^0 + \nu RT \ln a_{\pm} = \mu_2^0 + \nu RT \ln m\gamma, \quad (269)$$

m and γ being the mean molality and activity coefficient in the *saturated solution*. The chemical potential of the solute when the ions are in their standard states is μ_2^0 , and so the increase of free energy (ΔF^0) in transferring 1 mole of salt from the solid to the solution in the standard state is

$$\Delta F^0 = \mu_2^0 - \mu_s = -\nu RT \ln m\gamma. \quad (270)$$

It is thus possible to evaluate ΔF^0 from a determination of the solubility of the salt, which gives m in the saturated solution, and the activity coefficient in this solution. The heat content change ΔH^0 is equal to the heat of solution, per mole of solute, at high dilution, when the activity coefficients are almost unity. Since ΔF^0 and ΔH^0 are known, the corresponding ΔS^0 value, called the **entropy of solution**, can be obtained. The molality of a saturated solution of potassium chlorate, for example, is 0.715 at 25°, and the mean activity coefficient may be taken as 0.476, which is the value found in a potassium nitrate solution at the same concentration; the two salts have almost identical coefficients below 0.1 M, and so the assumption made is not likely to be seriously in error. The standard free energy of solution is thus

$$\Delta F^0 = -2RT \ln (0.476 \times 0.715) = 1277 \text{ cal.}$$

The heat of solution ΔH^0 , for a dilute solution, is 10,050 cal. at 25° c., and so the entropy of solution ΔS^0 is $(10,050 - 1,277)/298.2 = 29.4$ E.U.

From specific heat measurements the absolute entropy, based on the third law assumption (cf. p. 866), of solid potassium chlorate is 34.2 E.U. per mole; if this is added to the entropy of solution, the result, i.e., 63.6, is the sum of the "absolute" entropies of the potassium and chlorate ions in their standard states. Since $\bar{S}_{\text{K}^+}^0$ has been shown above to be 24.2, $\bar{S}_{\text{ClO}_3^-}^0$ is 39.4 E.U. per g. ion at 25° c.

By the use of the above and other methods the standard entropies recorded in Table 168 have been obtained (W. M. Latimer, 1938). One of the most important applications of the standard entropies is to the calculation of electrode potentials which cannot be measured directly; two examples will be considered. The heat change of the reaction

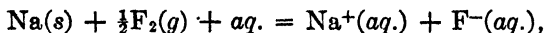


TABLE 168. STANDARD ENTROPIES OF IONS AT 25° C. IN CAL. DEG⁻¹ G. ION⁻¹

Al ⁺⁺⁺	-76	Ca ⁺⁺	+11.4	SO ₄ ⁻	+ 4.4
Fe ⁺⁺⁺	-63.3	Cd ⁺⁺	+16.4	NO ₃ ⁻	+35.0
Mg ⁺⁺	-31.6	Pb ⁺⁺	+ 3.9	I ⁻	+25.3
Fe ⁺⁺	-25.9	K ⁺	+24.2	Br ⁻	+19.7
Zn ⁺⁺	-25.7	Ag ⁺	+17.5	Cl ⁻	+13.5
Cu ⁺⁺	-26.5	Hg ₂ ⁺⁺	+17.7	F ⁻	- 2.3
Hg ⁺⁺	- 6.5	Na ⁺	+14.0	OH ⁻	- 2.5

that is the heat of formation of a dilute sodium fluoride solution from the elements, is $\Delta H^0 = -135,500$ cal. at 25° C. The entropy of Na(s) is 12.2 and that of $\frac{1}{2}\text{F}_2(g)$ is 24.0 E.U. per g. atom, and these together with the values for Na⁺ and F⁻ from Table 168 give

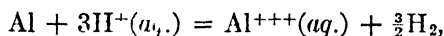
$$\Delta S^0 = \bar{S}_{\text{Na}^+}^0 + \bar{S}_{\text{F}^-}^0 - S_{\text{Na}} - \frac{1}{2}S_{\text{F}_2} = -24.5 \text{ E.U.}$$

for the same reaction. The standard free energy change accompanying the formation of the ions from their elements is, therefore,

$$\Delta F^0 = \Delta H^0 - T\Delta S^0 = -128,300 \text{ cal.}$$

Taking the free energies of the elements, as before, to be zero, this is the sum of the standard free energies of the sodium and fluoride ions, which is equal to $0.2390F(-E_{\text{Na}^+}^0 + E_{\text{F}^-}^0)$ cal. Since $E_{\text{Na}^+}^0$ is +2.71 volts, it follows at once that $E_{\text{F}^-}^0$ is -2.85 volts at 25° C.

The potential of the aluminum electrode can be calculated in an analogous manner. The heat of solution of aluminum in dilute acid, i.e.,



gives $\Delta H^0 = -126,300$ cal. at 25°. The entropy of Al is 6.7, and since that of H₂ is 31.2 and $\bar{S}_{\text{H}^+}^0$ is assumed to be zero, it follows that

$$\Delta S^0 = \bar{S}_{\text{Al}^{+++}}^0 + \frac{3}{2}S_{\text{H}_2} - S_{\text{Al}} - 3\bar{S}_{\text{H}^+}^0 = -36 \text{ E.U.}$$

From this and ΔH^0 it is found that ΔF^0 is -115,500 cal., which is equivalent to the free energy of formation of Al⁺⁺⁺ ions, since that for H⁺ ions is zero, by convention. It follows, therefore, that $-0.2390 \times 3FE_{\text{Al}^{+++}}^0$ is -115,500 cal., so that $E_{\text{Al}^{+++}}^0$ is +1.67 volts at 25° C.³⁶

ELECTROLYSIS AND POLARIZATION

Electrolytic Polarization.—At a reversible electrode, which is in a state of equilibrium, the discharge of ions and their formation take place at the same rate (p. 948), and there is no net flow of current. If, however, as the result of the application of an external E.M.F. there is an actual passage of current, the electrode is disturbed from its equilibrium condition; this disturbance of equilibrium associated with the flow of current is called **electrolytic polarization**, and the disturbed electrode is said to be **polarized**. According to modern views, polarization results from the slowness of one or more of the processes occurring at the electrode during the discharge or the formation of an ion; the type of polarization depends

on the nature of the slow process. Polarized electrodes are often said to behave **irreversibly**, the irreversibility arising because of the slowness of one or other of the stages in the electrode process. A simple type of polarization, which is due essentially to the slowness of the diffusion of ions in solution, results from changes of concentration taking place in the vicinity of an electrode during electrolysis; this is known as **concentration polarization**. For example, when a metal anode dissolves there is an increase of ionic concentration in its immediate neighborhood, while there is a decrease of concentration near the cathode, where ions are being discharged. A substance which tends to reduce the extent of polarization is called a **depolarizer**, e.g., an oxidizing agent in the vicinity of a cathode when there is a tendency for hydrogen to be evolved.

Decomposition Voltage.—If two small sheets of smooth platinum are immersed in dilute sulfuric acid and a gradually increasing E.M.F. is applied by means of an external battery, so as to electrolyze the solution, a galvanometer in the circuit will show that the current increases in the manner represented by the curve in Fig. 243. It appears that an appreciable voltage must be applied

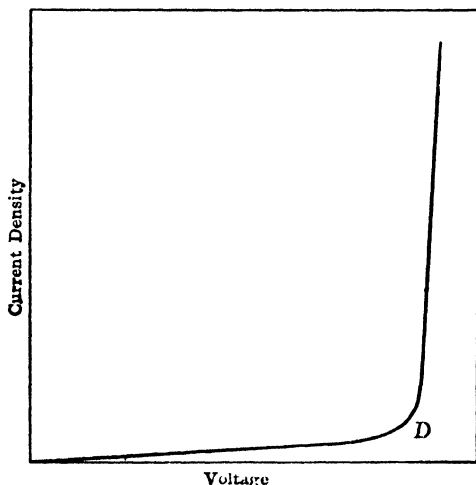


FIG. 243. Decomposition voltage

before current can flow freely through the cell. The E.M.F. applied at the point *D* at which steady electrolysis commences is called the **decomposition voltage** of the particular solution, with the given electrode material. It is at this point that, in general, the steady evolution of hydrogen and oxygen gas bubbles is observed. Attention should be drawn to the fact that the nature of the current-voltage curve in Fig. 243 is such that *D* cannot be identified precisely; in any event, it is now believed not to have any exact theoretical significance.

The apparent decomposition voltage is nevertheless of interest, for it gives an approximate indication of the minimum E.M.F. which must be applied to a particular solution, in the absence of depolarizers, in order that currents of appreciable magnitude shall be able to pass. It represents the sum of the potentials which must be attained by the two electrodes before the rates of the respective ion discharge processes are appreciably greater than the reverse reactions.

The decomposition voltages of aqueous solutions of acids and bases were studied systematically by M. Le Blanc (1891–93); some of his re-

TABLE 169. DECOMPOSITION VOLTAGES OF ACIDS AND BASES IN AQUEOUS SOLUTION

Acid	Volts	Base	Volts
N HClO_4	1.65	N KOH	1.67
$\text{N CCl}_3\text{CO}_2\text{H}$	1.66	$\frac{1}{2}\text{N NH}_2(\text{CH}_2)_2\text{OH}$	1.68
$\text{N H}_2\text{SO}_4$	1.67	N NaOH	1.69
N HNO_3	1.69	$\text{N NH}_4\text{OH}$	1.74
$\text{N H}_3\text{PO}_4$	1.70	$\frac{1}{3}\text{N N}(\text{CH}_3)_4\text{OH}$	1.74

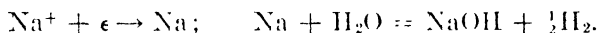
sults, obtained with smooth platinum electrodes, are quoted in Table 169. It is seen that the values are all in the vicinity of 1.7 volts, and hence Le Blanc concluded that the same electrolytic process occurred at the electrodes in each case. The only process which could be common to all aqueous solutions of acids and bases is the discharge of hydrogen ions at the cathode and of hydroxyl ions at the anode, to form hydrogen and oxygen, respectively. It appears, therefore, that although the current is actually *carried* by all the kinds of ions in the electrolyte, only certain cations, not necessarily those present in excess, are actually discharged at the electrodes. Thus, in sodium hydroxide solution, for example, the current is carried toward the cathode almost exclusively by sodium ions, but it is nevertheless the hydrogen ions that are discharged, at least at a platinum cathode.

Discharge or Deposition Potentials.—The conditions determining whether a particular ion can be discharged or not can be found by measuring the separate potential at each electrode during electrolysis. Instead of observing the voltage across the whole cell, the potentials of the electrodes are obtained individually by connecting each with a reference, e.g., calomel, electrode. The variation of electrode potential with current is depicted by a curve similar to that in Fig. 243, the point *D* now corresponding to the **discharge potential**, or **deposition potential**, at the particular electrode. In this way M. Le Blanc (1893) found that, in general, the discharge potential of an ion, e.g., cadmium, silver, zinc and copper, was approximately equal to the static, or equilibrium, potential in the same solution. This result is to be expected for any electrode which behaves in an almost reversible manner, i.e., one for which there is no appreciable retardation of the electrode processes. A small increase in potential over the equilibrium value will then make the discharge of ions take place at a considerable rate. The majority of metals behave in this manner, the discharge potential, which is in any case an approximate quantity, being close to the reversible potential in the same solution.

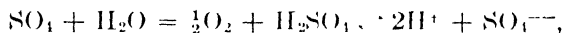
If an electrolyte contains a number of different ionic species then, provided there are no disturbing factors, each ionic discharge will occur as the appropriate potential is reached. Since the process at a cathode involves the taking up of electrons, i.e., reduction in the most general sense, cathodic processes will take place in the order of decreasing reduc-

tion potentials,* that is, in the order of increasing oxidation potentials. Similarly, anodic processes will occur in the order of decreasing oxidation potentials. In an acid solution of copper sulfate, for example, there are present cupric and hydrogen ions; the oxidation potentials are -0.34 and ± 0.0 volt, respectively (Table 156), and so, in accordance with the rule given above, cupric ions will be discharged first at the cathode. If the anode in the same solution is of copper, then three anodic processes are possible, viz., (a) solution of the copper to form copper ions, (b) discharge of sulfate ions, (c) discharge of hydroxyl ions. The oxidation potentials for these processes in the given solution are about -0.34 , -2.0 and -1.2 volt, respectively; hence, the copper anode will dissolve in preference to the other processes. With a platinum anode, however, the corresponding oxidation potential has a large negative value, and so there will be a discharge of hydroxyl ions accompanied by the evolution of oxygen.

It is frequently stated that in the electrolysis of an alkali hydroxide solution, the alkali metal is first discharged, and that the free atom attacks the water to give hydrogen; thus,



Since the sodium hydroxide is completely ionized, and hence is equivalent to Na^+ and OH^- , the mechanism postulated is equivalent to saying that the sodium ion is not discharged and that the hydrogen comes from the water. Similarly, the view that SO_4^{--} ions are discharged at the anode in the electrolysis of aqueous sulfuric acid solution and that the SO_4 radicals attack water, does not take into account the fact that the latter reaction, viz.,



proves that SO_4^{--} ions are reluctant to be discharged. Le Blanc and others have expressed the opinion that hydrogen and hydroxyl ions are always discharged directly, at cathode and anode, respectively, in solutions of both acids and bases. In spite of the small concentration of hydrogen ions, for example, in a solution of sodium hydroxide, these ions are discharged in preference to sodium ions. Similarly, in aqueous sulfuric acid it is believed that hydroxyl ions, and not sulfate ions, are discharged, although the former are present in a concentration of about 10^{-14} g. ion per liter. If the same process is occurring at anode and cathode, irrespective of the nature of the electrolyte, the constancy of the decomposition voltage (Table 169) can be readily understood. A further clarification of this problem will appear later in connection with the treatment of overvoltage (p. 1033). Whatever the mechanism of the electrolysis of aqueous solutions, it is important to remember that the ions which carry

* The greater the reduction potential the greater the tendency for the process $\text{M}^+ + e \rightarrow \text{M}$ to take place, and hence the more readily will the discharge of M^+ ions occur.

the current to the electrodes are those present in excess, but the oxidation or reduction potentials determine the order in which the ions are discharged.

When solutions of halogen acids or of halides are electrolyzed, halogen ions, with the exception of fluorine, are first discharged; this process is, however, soon followed by the evolution of oxygen, especially in chloride solutions. The results are in agreement with expectation, for, in acid solutions especially, the oxidation potentials of the halogen ions are larger than that required for oxygen evolution.

In view of the above discussion it may appear surprising that the decomposition voltages of solutions of alkali and alkaline-earth salts of strong acids, e.g., sodium nitrate, potassium sulfate, calcium nitrate, etc., are about 2.2 volts; if direct hydrogen and hydroxyl ion discharge occurs, as suggested, the voltage should be 1.7. The explanation is that the discharge of these ions leaves an excess of hydroxyl and hydrogen ions in the vicinity of cathode and anode, and since the solutions are not buffered they become alkaline and acid, respectively. Hydrogen is thus liberated from an alkaline solution and oxygen from an acid electrolyte, and this must introduce an additional voltage. As seen on p. 1006 a solution of an acid or base has much greater buffer capacity than a salt, and so the changes in ionic concentration have little effect in the electrolysis of the former.

Overvoltage.—The theoretical voltage required for the reversible decomposition of an aqueous solution of an acid or a base is equal to the E.M.F. of the reversible cell with hydrogen and oxygen gases at 1 atm. pressure; this is known to be about 1.23 volts at ordinary temperatures, irrespective of the nature of the electrolyte (cf. p. 946), provided of course that it is the same throughout the cell. The fact that the observed decomposition voltage is much larger than the theoretical (reversible) value means that the flow of an appreciable current, requiring the discharge of hydrogen and hydroxyl ions at appreciable rates, is accompanied by a relatively large polarization. This polarization evidently depends on the nature of the electrodes, for with a lead anode and cathode the decomposition of dilute sulfuric acid solution requires about 2.2 volts, instead of 1.7 volts for two platinum electrodes. A detailed study of this problem was made by W. A. Caspari (1899), who observed the potentials at which visible evolution of hydrogen and oxygen gases occurred at cathodes and anodes, respectively, of a number of different metals. The difference between the potential of the electrode when gas evolution was actually observed and the theoretical (reversible) value for the same solution was called the **overvoltage**.* The discussion here will be restricted to hydrogen, i.e., cathodic, overvoltage, and the phenomena at the anode will be considered later. A selection of the hydrogen overvoltages determined by Caspari are recorded in Table 170; it is seen that only for

* The original word used by Caspari was "Überspannung"; it is translated as "over-voltage" or "overpotential."

TABLE 170. CATHODIC (BUBBLE) OVERVOLTAGES IN DILUTE SULFURIC ACID

Platinized Platinum	0.005 volt	Copper	0.23 volt
Gold	0.02	Cadmium	0.48
Iron (in NaOH)	0.08	Tin	0.53
Smooth platinum	0.09	Lead	0.64
Silver	0.15	Zinc	0.70
Nickel	0.21	Mercury	0.78

a platinized platinum cathode does the evolution of hydrogen occur in the close vicinity of the reversible potential. With a mercury cathode the cathodic potential must exceed the reversible value by about 0.78 volt before appreciable hydrogen evolution can be observed in acid solution.

The concept of overvoltage was extended by J. Tafel (1905); he measured the excess potential, over the reversible value, at a cathode at which hydrogen was being evolved at a *definite current density*.^{*} It has become the practice to refer to this quantity by the general term overvoltage, the c.d. being stated. It is not certain that the quantity measured by Caspari, which corresponds roughly to the potential at which appreciable current begins to flow (cf. point *D* on Fig. 243), has any fundamental theoretical significance. Nevertheless, the results are of considerable practical interest in connection with the possibility of hydrogen evolution in various electrolytic processes. For this reason the point at which gas bubble formation commences is of some importance, and the results given in Table 170 are referred to as "bubble overvoltages."

The determination of the bubble overvoltage requires careful observation of the electrode to find the point at which bubble evolution commences as the applied E.M.F. is gradually increased, or that at which visible evolution ceases on decreasing the E.M.F. The potential of the cathode is then measured against a reference electrode, and compared with that of a reversible, e.g., hydrogen gas, electrode in the same solution. In the measurement of overvoltage at a definite c.d., the cathode is allowed a short time to come to equilibrium, with the given current flowing, and the potential is then observed. If the current strength is fairly high, e.g., greater than about 0.01 amp., an appreciable error due to the fall of potential at the electrode surface may be involved. This can be avoided by using small currents, preferably in a solution saturated with hydrogen, or by determining the back E.M.F. of the electrode at various short intervals after switching off the polarizing current and extrapolating to zero time. The results of overvoltage measurements are not exactly reproducible, because various factors, e.g., electrode surface and traces of impurities, which are not easy to control have an important influence; nevertheless, certain general tendencies are quite clear and these will be considered.³⁷

^{*} Current density, abbreviated to c.d., is defined as the current per unit area of electrode surface; it is generally expressed as amp. or milliamp. per sq. cm. or per sq. dm.

Results: Bubble Overvoltage.—The data in Table 170 may be taken as approximately correct, although somewhat different values have been reported. The effect of surface is seen by comparing the figures for platinized and smooth platinum; a similar difference is found for other metals, a roughened surface having invariably a lower overvoltage than a smooth electrode of the same metal. That surface forces may be of importance is suggested by the appreciable reduction of bubble overvoltage accompanying the addition of ethyl and methyl alcohols, acetic acid or acetone to the aqueous electrolyte; these substances all bring about a lowering of surface tension. Strongly adsorbed substances, such as amyl alcohol, result in a marked decrease of surface tension, but they have a tendency to raise the overvoltage. The bubble overvoltage of a given cathode appears to be independent of the pH of the solution over a wide range, viz., 0 to 12, but it decreases as the temperature is raised.

Overvoltage at Definite C.D.—Provided the conditions are such that the hydrogen does not diffuse away rapidly from the cathode, the overvoltage (ω) increases in a logarithmic manner with the c.d., (I); thus,

$$\omega = a + b \log I, \quad (270)$$

where a and b are constants. At ordinary temperatures the factor b has an approximately constant value of about 0.12 for a number of metals, e.g., copper, silver, gold, nickel and mercury. At platinum and palladium cathodes, however, b may be as low as 0.05, and at lead in N acetic acid, and tantalum in sulfuric acid, it may be greater than 0.2.

The overvoltage at a definite c.d. decreases if the electrode surface is roughened; this is due partly, but not entirely, to an increase in the effective area and the consequent decrease in the actual c.d. As the temperature is raised overvoltage diminishes; for most electrodes the temperature coefficient is about -0.002 volt per degree, but this is only an approximate figure, as it varies with the nature of the metal and its actual overvoltage. Organic liquids, such as methyl and ethyl alcohols, which lower the surface tension of aqueous solutions, lower the overvoltage, but colloidal materials raise the latter although they may diminish the surface tension. The result of changing the hydrogen ion concentration is not certain, but it appears that at low c.d.'s the overvoltage is independent of the pH, at least in acid solutions.

The rate of increase of potential in the attainment of overvoltage, in a solution saturated with hydrogen, is found to be constant; that is to say, the potential is a linear function of the quantity of electricity passed. Only when the electrode is approaching its final potential for the given c.d. does the rate fall off. The result implies that during the early stages of the growth of the overvoltage potential, the whole of the current goes to charge up a double layer consisting of electrons, or negative ions, on the electrode side and of hydrogen ions on the solution side. The measured cathode potential at any instant is the actual potential across this double layer which behaves as a condenser; the capacity is probably of the order

of 10^{-5} coulomb per volt per sq. cm. of actual electrode surface, irrespective of the nature of the metal. From this result it appears that very minute amounts of current, much less than would be required to produce a hydrogen layer one molecule thick, are able to establish the final overvoltage condition (F. P. Bowden and E. K. Rideal, 1928). It must be remembered, however, that these conclusions apply to a solution, and electrode, previously saturated with hydrogen.³⁸

Theories of Overvoltage.—When an electrode is at its reversible potential ions are being discharged from the solution at the same rate as they return to the solution from the electrode (cf. p. 948); there is then no resultant flow of current. If current is to pass through the solution, however, one of these two processes must occur more rapidly than the other: for example, at a cathode the hydrogen ions must be discharged to form hydrogen at a greater rate than the hydrogen returns to the solution as ions. If any of the stages involved in the discharge of hydrogen ions occurs slowly, an appreciable potential, in excess of the reversible value, must be applied to the electrode. The cause of overvoltage is, therefore, the relative slowness of one or more steps between hydrogen ions in solution, on the one hand, and hydrogen gas on the electrode, on the other hand. The processes occurring at the cathode in an aqueous solution may be considered as follows: (1) transport of H_3O^+ ions to the electrode layer; (2) transfer of these ions, or possibly protons, to the electrode; (3) discharge of the ions or protons by electrons; (4) combination of the resulting hydrogen atoms to form molecules; and (5) evolution of the hydrogen molecules in bubbles of gas. It is improbable that (1) is rate-determining because the overvoltage depends on the nature of the metal forming the cathode, and although (5) undoubtedly plays some part, there is reason to believe that it is small. The view that the formation of molecular hydrogen from atoms, i.e., stage (4), was a slow process was proposed by J. Tafel (1900) and received support from many others; according to this theory overvoltage is due to the high potential brought about by hydrogen atoms. It can be calculated from the free energy of the reaction $2H \rightarrow H_2$, that the standard potential of atomic hydrogen at 1 atm. pressure should be about 1.9 volts, and hence a pressure of 10^{-10} atm. would be sufficient to account for the highest known overvoltages, viz., about 1.3 volts. The variation of overvoltage from one metal to another may be ascribed to their differing catalytic effect on the rate of combination of hydrogen atoms; the low overvoltage metals, e.g., platinum, palladium and nickel, should be good catalysts and those of high overvoltage, e.g., lead and mercury, may be expected to be poor catalysts. Direct experimental studies with atomic hydrogen bear out this anticipation. Further, low overvoltage metals are the best catalysts for hydrogenation reactions, in which it is probable that $H_2 \rightarrow 2H$ is a preliminary step. If the metals catalyze this process they should also catalyze the reverse reaction (p. 1128), and so facilitate the combination of atomic hydrogen to form molecules. The decrease of overvoltage accompanying an increase of temperature is attributed to the increased velocity of this reaction.

The chief objection to the theory outlined is that if the rate of combination of hydrogen atoms is proportional to the square of their concentration, as would be the case if the reaction were homogeneous, the variation of overvoltage with current should be given by

$$\omega = a + (RT/2F) \ln I = a + (2.30RT/2F) \log I, \quad (271)$$

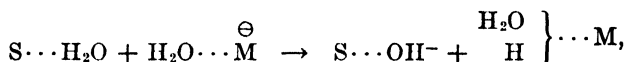
where a is a constant. This is of the same form as (270), but the constant b in the latter equation, representing the slope of the straight line plot of ω against $\log I$, would be $2.30RT/2F$, which is 0.03 at 25°, instead of the experimental value of 0.12 found in many cases.

The view that the discharge of hydrogen ions, i.e., stage (3), occurs slowly is implicit in the theory of A. Smits (1922), but it was put forward more precisely by T. Erdey-Grúz and M. Volmer (1930). If the ions brought up to the electrode by the current are not discharged immediately, but form with the electrons on the cathode a double layer, which may be regarded as constituting two plates of a condenser, then the difference of potential across it gives the measured electrode potential. If during the process of building up this double layer, by ions and electrons, practically no ions are discharged, the potential across the layer will be proportional to the quantity of electricity passed, as found experimentally. It is postulated that for an electron to pass across the double layer and so discharge a hydrogen ion, the ion-electron doublet must acquire a certain activation energy (cf. p. 1089). As the double-layer potential increases the energy of the doublet is increased, and the requisite activation energy is correspondingly lowered; the rate of transfer of electrons, and hence the rate of discharge of hydrogen ions, is consequently increased. When the double-layer potential reaches the overvoltage value the discharge of hydrogen ions takes place at the rate required by the current passing. On the basis of these views it is possible to derive an equation of the form of (270), where b is equivalent to $2.3RT/\alpha F$; the factor α is less than unity, and if it is 0.5, as is quite reasonable, the value of $2.3RT/\alpha F$ is $2.3 \times 2RT/F$, i.e., 0.12 at 25° C., in agreement with experiment.

Another theory, in a sense involving the concept of slow neutralization of the charge on the hydrogen ion, has been developed by R. W. Gurney (1931) using the methods of quantum mechanics. It is supposed that an energy barrier exists between the electrons in a metal and an ion approaching it, and according to classical mechanics the electron cannot cross the barrier to discharge the ion; quantum mechanics, however, allows of a definite probability of "leakage," through the barrier, of an electron from the cathode to an unoccupied level of the same energy in the ion to be discharged. If the highest occupied energy level in the cathode is well below that of the (hydrated) hydrogen ion in solution neutralization of the latter cannot occur; that is to say, the ion cannot be discharged. By applying a negative potential to the electrode, however, the energy of the electrons is raised until they are at a level higher than that of the ions in solution. In order to calculate the strength of the current which is able to pass at a particular potential, the latter giving the overvoltage at that current, it is necessary to determine the rate of transfer of electrons through the barrier, allowance being made for the facts that not all the electrons in the metal are at the same energy level and that the hydrated ions do not all have the same energy because different vibrational levels are possible. In this way an equation of the correct form relating overvoltage to current strength can be obtained, and by making a fairly reasonable assumption the value of b in (270) is found to be in agreement with experiment. A weakness of this, and the previous, theory is that it is not clear what specific property of the metal is responsible for its overvoltage, and the catalytic properties of low overvoltage metals are given no adequate interpretation.

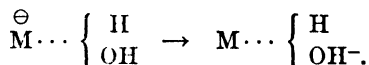
It is not altogether easy to accept the view that the discharge of an ion by an electron is a slow process, and H. Eyring, S. Glasstone and K. J. Laidler

(1939) have developed a theory based on the idea that stage (2), i.e., the transfer of a proton, is the rate-determining step and hence the cause of overvoltage. It is suggested that, in the absence of strongly adsorbed substances, an electrode in an aqueous solution becomes covered to a large extent with a unimolecular layer (see Chapter XIV) of water molecules. Adjacent to this layer there is another layer of water molecules associated with the bulk of the electrolyte, and the transfer of a proton from a molecule of water on the solution side to one on the electrode is regarded as the slow process responsible for overvoltage. The proton is immediately discharged by an electron to form a hydrogen atom and a hydroxyl ion is left in solution. The essential discharge mechanism is thus



where the symbol S refers to the solution and M to the electrode; the \ominus sign on M is used to indicate an electron which is always available at the cathode. Not only are the experimental facts entirely consistent with this theory treated as a rate process requiring a definite energy of activation, but it accounts satisfactorily for the fact that the constant b in (270) is approximately $2.3 \times 2RT/F$. If the electrode material is one which adsorbs hydrogen readily, e.g., platinum, palladium, nickel or copper, it will form relatively strong M—H bonds, and the passage of a proton from the solution layer to the electrode layer will occur more readily than if the metal were one such as mercury or lead which forms only feeble M—H bonds. The former may thus be expected to have relatively low overvoltages, as is actually the case. Metals forming strong M—H bonds will also be good catalysts for the combination of atomic hydrogen to form molecules, but the parallelism between this property and overvoltage, mentioned above, does not necessarily mean that overvoltage is due to atomic hydrogen. The low overvoltage and the catalytic properties of platinum, palladium, nickel, etc., are both due to the same fundamental cause, namely the formation of strong M—H bonds.

The linear growth of overvoltage, referred to on p. 1019, can be readily explained by supposing that water molecules exist on the electrode surface as H and OH, and that while the overvoltage potential is building up the OH radical unites with an electron to form an OH^- ion on the surface; thus,



This is the only cathodic process before the overvoltage is attained, and so the number of OH^- ions produced will be directly proportional to the quantity of electricity passed. If the OH^- ions in the electrode layer are balanced by an equal number of positive ions in the solution layer, brought up by the current, the double-layer potential v will increase with time in a linear manner, as found in practice.

It may be mentioned, in conclusion, that when the overvoltage is very low, e.g., at platinized platinum, the constant b is approximately $2.3RT/2F$; under these conditions it is quite possible that the combination of hydrogen atoms is the slow process responsible for overvoltage (L. P. Hammett, 1933).³⁹

Cathodic Deposition of Metals.—Apart from iron, cobalt and nickel, the common metals commence to be deposited from solution near their

reversible potentials, and at very small c.d.'s the discharge of metal cations may be regarded as an almost reversible process occurring with little overvoltage. At increasing c.d.'s the electrode potential becomes slightly more negative, partly as a result of concentration polarization and partly on account of the slowness of one or other of the stages which must be involved between the initial discharge of the ion and the final orientation of the atom in the crystalline deposit. Neutralization of the ion probably occurs rapidly, but there may be some delay in the growth of the crystal nuclei.

Polarization of quite a different type occurs in the deposition of iron, cobalt and nickel. Even at extremely small c.d.'s deposition does not take place unless the potential is distinctly in excess of the reversible value; in other words, there is a definite overvoltage for the deposition of these metals. At 20° c. the overvoltages for the commencement of deposition are 0.2 to 0.3 volt, being least for iron and most for nickel, but the irreversibility decreases as the temperature is raised. Iron deposits reversibly at temperatures above 70°, but cobalt and nickel have deposition overvoltages of about 0.05 volt even at 100°. It was thought at one time that the polarization accompanying the deposition of the iron-group metals was connected with the hydrogen which is liberated simultaneously, but this view is probably incorrect as the potentials required for the liberation of the metals are not affected by changes in the pH of the electrolyte. The metal overvoltage is thus independent of the extent of the accompanying discharge of hydrogen ions. The most plausible explanation of the observations is that iron, cobalt and nickel are initially deposited from solution in an active or unstable state. The overvoltage would then be ascribed to the relative slowness of the conversion of the active to the normal form of the metal on the cathode. Increase of temperature should increase the rate of conversion of the unstable to the stable form, so that the deposition potential should approach the theoretical figure (S. Glasstone, 1926).

Depolarization of Metal Deposition.—If the metal being deposited can form a solution in the cathode or, better, form a compound with it which dissolves in the cathode material, then the partial molar free energy of the metal is decreased and deposition can occur at a potential less cathodic than the reversible value.* A striking example of this type of depolarization is the deposition of sodium, or other alkali metal, on a mercury cathode, from a neutral or alkaline solution. The reversible deposition potential of sodium from a solution containing about 1 g. equiv. per liter of a sodium salt is about -2.7 volts (cf. Table 156), but as a

* In connection with cathodic deposition potentials *measured during electrolysis*, the convention as to sign is that for equilibrium reduction potentials. A less (or more) cathodic potential means a less (or more) negative potential on this basis, and means that the process can take place more (or less) easily. The order of the discharge of cations is the order of increasing cathodic, i.e., negative, deposition potentials.

consequence of the formation of compounds, i.e., amalgams, which are soluble in mercury, appreciable discharge of sodium ions occurs at -1.2 volts. In a neutral solution of a sodium salt, i.e., $pH\ 7$, the theoretical discharge potential for hydrogen ions is -0.4 volt, but at a mercury cathode there is an overvoltage of about 0.8 ; hence the evolution of hydrogen does not commence until the cathode attains a potential of -1.2 volts. It is clear, therefore, that owing to the high hydrogen overvoltage at a mercury cathode, and the marked depolarizing effect of the latter on the discharge of sodium ions, the simultaneous deposition of sodium and hydrogen can occur. It is thus possible to prepare dilute alkali amalgams by the electrolysis of aqueous alkali chloride solutions, and this fact has been utilized in the electrolytic processes for the manufacture of sodium hydroxide.

Separation of Metals by Electrolysis.—The electrolytic method for the quantitative separation of metals was employed by Wolcott Gibbs as far back as 1864, and the importance of adjusting the E.M.F. was realized by H. Kiliani (1884), but it was not until Le Blanc reported his work on deposition potentials (p. 1015) that the theoretical basis of the separation became clear (H. F. Freudenberg, 1893). As explained above, apart from polarization effects, a metal commences to be deposited from a solution when the cathode potential is made slightly more negative than the reversible value in that solution. If a solution contains 1 mole per liter of each of a simple salt of silver, copper and cadmium, for example, deposition of the metals will commence at $+0.79$, $+0.32$ and -0.40 volt, respectively.* Hence by regulating the applied E.M.F., and thus the cathode potential, the silver may be deposited virtually completely before copper commences to deposit, and the copper may similarly be separated from the cadmium. These separations are, of course, contingent on the metals not forming compounds or solid solutions which cause them to be deposited simultaneously. Since the concentration of the ions of deposited metal decreases during electrolysis, its deposition potential becomes steadily more negative; thus if the ionic concentration is reduced to 0.1 per cent of its original value, then it can be seen from (88) that the potential will be changed by 3×0.029 i.e., 0.087 , volt for a bivalent metal, and 3×0.058 , i.e., 0.174 , volt for a univalent metal. The initial deposition potentials of two metals should thus differ by at least 0.2 volt so that the concentration of one may be reduced to an analytically negligible amount before the other commences to deposit. If the potentials do not normally differ by this amount it may be possible to make them do so by converting one of the cations into a complex anion. In their simple salt solutions the potentials of copper and bismuth are so close that a satisfactory electrolytic separation is not possible. If an alkali cyanide solution is added, however, the copper ions form a complex cuprocyanide, and since the concentration of simple cupric ions is thereby very markedly diminished, the copper potential becomes much more negative, viz., about -1.0 volt.

* See preceding footnote (p. 1023)

The bismuth ions form no such complex and the cathodic potential remains at about $+0.2$ volt, so that a complete separation from copper is possible.

When a metal does not deposit reversibly, it is necessary to make allowance for the deposition overvoltage when considering its separation from other metals; sometimes it is necessary to raise the temperature to avoid the difficulty which may arise. If an ammoniacal solution containing zinc and nickel is electrolyzed at 20°C ., both metals deposit simultaneously; at 90° , however, complete separation of the nickel, which is deposited first, is possible. At the latter temperature the deposition overvoltage for nickel is nearly 0.3 volt less than at 20° , but the zinc deposits reversibly at both temperatures, so that at 90° the two potentials are sufficiently different to permit of a satisfactory separation.

It is of interest to note that, at first sight, it would appear to be impossible to deposit zinc from a simple salt solution containing acid, whereas in actual fact such deposition does occur. The cathode potential of zinc in a molar solution of its sulfate, for example, is -0.76 volt, but the hydrogen potential in an acid solution is about ± 0.0 volt; hydrogen evolutions should thus occur in preference. At a zinc cathode, however, the overvoltage for hydrogen ion discharge is very high, over 0.7 volt, and so potentials of the same order are required for the separation of zinc and hydrogen; consequently, simultaneous deposition of the two elements is observed.

Simultaneous Deposition of Metals.—It is sometimes desirable, e.g., in the preparation of alloys, to deposit two metals simultaneously by electrolysis; this can be done if their deposition potentials are similar. The metals copper and zinc are deposited separately from a solution containing their mixed sulfates, because the requisite potentials are $+0.34$ and -0.76 volt, respectively; if, however, an excess of alkali cyanide is added to the solution, both metals form complex cyanides from which deposition takes place at about -1.0 volt in each case. The cuprocyanide complex is much more stable than that formed by zinc, and so the concentration of free copper ions is reduced to a much greater extent than zinc ions; the potentials of the two electrodes are thus brought close together. By electrolysis of a solution of copper and zinc salts containing excess of cyanide it is possible to prepare alloys of the two metals. Mixtures containing cadmium, with either copper or zinc or both, can be deposited in a similar manner.

Physical Nature of Electrodeposited Metals.—In view of the industrial importance of the electrodeposition of metals, the influence of various factors on the physical appearance of the deposits has been the subject of much investigation. The position is complicated, but there are a few general rules which appear to be applicable in many instances. At low c.d.'s the deposits are coarsely crystalline, because the rate of formation of new nuclei by the discharge of ions is less than the rate of growth of those already present: at higher c.d.'s the former is increased and the deposits become fine-grained. At

very high c.d.'s, however, nodules, protruding crystals or "trees" form, chiefly because of the depletion of the concentration of electrolyte near the cathode. If the concentration is kept high, then the roughening effect may be considerably delayed; for this and other reasons the solutions employed in electrolysis are agitated. Increase of temperature has two opposing effects: in the first place, it favors diffusion and so tends to produce a uniform, fine deposit but, on the other hand, it aids the rate of crystal growth. The first factor generally predominates at moderate temperatures, and the second becomes more effective as the temperature is raised. The nature of the electrolyte plays an important part in determining the form of the deposit, but the reason for this is not quite clear. Silver is obtained as coarse crystals from the nitrate solution, but electrolysis of the complex cyanide gives the familiar smooth deposit of silver plate. Rough deposits of lead are also obtained from nitrate solutions, but they are fine grained from silicofluoride or borofluoride solutions. It appears, as a rough generalization, that better deposits are obtained from complex ions than from simple ions of the same metal. The presence of very small amounts of colloidal matter, or of certain organic compounds, frequently favors the production of smooth, fine-grained, microcrystalline electrodeposits. A wide variety of materials, some of which are employed in technical practice, are effective in this connection, e.g., gelatine, peptone, rubber, dextrin, glue, casein, glucosides, alkaloids, camphor, etc. It is not improbable that the good deposits obtained from solutions of complex ions are often due to small amounts of colloidal matter formed at the cathode. The behavior of the colloidal and other agents is no doubt connected with the fact that they are adsorbed on the crystal nuclei and so prevent their growth.⁴⁰

Polarization due to Slowness of Diffusion.—The various forms of polarization already considered were due to one of the stages in the ionic discharge process being slow; as stated earlier (p. 1014), a form of polarization, viz., concentration polarization, is to be attributed to the slowness of diffusion. The decrease of concentration in the vicinity of a cathode causes the potential to become more negative, while the increase near the anode results in a more positive potential. Since the effect of concentration on electrode potential is not very large (cf. p. 947), concentration polarization is usually small in magnitude, although certain unusual cases of large polarization are known. Concentration polarization can be diminished by increasing the rate of diffusion of the electrolyte, e.g., by vigorous agitation, by increasing the concentration of the solution, or by raising the temperature, but it cannot be entirely eliminated. It appears, from many observations concerned with electrolysis and other phenomena, that during electrolysis there is always immediately adjacent to the electrode a thin layer of solution in which there is a concentration gradient from that of the bulk of the electrolyte to the smaller or larger (for cathode or anode, respectively) value at the electrode. This is called the **diffusion layer**, and its apparent thickness, which is about 0.03 cm. in a still solution at ordinary temperatures, decreases with increasing temperature and agitation of the electrolyte. The actual concentration of the latter has only a minor effect on the thickness of the diffusion layer.

Before an ion can be discharged at the cathode it must traverse the

diffusion layer, from the bulk of the electrolyte to the electrode. There is a maximum rate at which this diffusion can occur, and for a given set of conditions it is proportional to the concentration of the ions in the solution. If the current is increased to a point at which the rate of diffusion is insufficient to keep pace with the discharge of the ions at the cathode, the potential will change to a more negative value where another process, requiring other ions, can occur. The current at which the potential changes in this manner represents the maximum rate of diffusion of the ions under the particular conditions. It is called the **limiting current** or **diffusion current** and it is proportional, as may be expected, to the concentration in the solution of the ions being discharged. This fact has been utilized for analytical purposes in the polarograph, described below, and in other similar ways.

The Mercury Dropping Cathode.—A special technique for the study of cathodic phenomena, viz., deposition of hydrogen or metals and electrolytic reduction, has been devised by J. Heyrovsky (1923 *et seq.*). The cathode is of mercury which is being continuously renewed by falling in drops from the end of a capillary tube. The anode is a layer of mercury at the bottom of the vessel; since it has a large area and the currents passed are generally very small, viz., about 10^{-6} amp., its potential remains almost constant and hence variations in E.M.F. of the cell are due almost entirely to changes in the cathode potential. The value of the latter can be measured directly by comparison with a standard electrode, but in the self-registering apparatus, known as the **polarograph**, the E.M.F. applied between the two mercury electrodes is steadily increased, and the corresponding value of the current is recorded photographically, or in other ways. The discharge of an ion, or the occurrence of a new process at the cathode, is indicated by a sudden rise of current with only a small increase of the E.M.F. across the cell. The polarograph has been used in many investigations of cathodic phenomena, including studies of deposition potentials of ions, hydrogen ion concentrations, and valence changes in electrolytic reduction. The results obtained are not always the same as those given by a stationary cathode, and it is possible that special factors are operative at the dropping mercury cathode; this makes it difficult to interpret some of the data obtained.

A particularly valuable use of the polarograph is for the analysis of a great variety of solutions as dilute a 10^{-6} molar; it can be adapted to almost any molecular or ionic species that can be reduced or discharged at the mercury cathode. The magnitude of the current which can pass at a particular potential is proportional to the concentration of the species reduced at that potential, in accordance with the discussion in the preceding section.⁴¹

Irreversible Anodic Phenomena.—Most metals dissolve as anodes when the potential is just slightly more positive than the reversible value in the particular solution used as electrolyte; as the c.d. is raised the potential increases somewhat because of concentration polarization, but the extent is not large. Anodes of iron, cobalt and nickel, however, exhibit a marked polarization which can only be ascribed to an irreversibility in the ionization process; this is not altogether surprising for it has been seen (p. 1023) that the reaction in the opposite direction is also irreversible.

In spite of the marked polarization, an anode of one of these metals still dissolves quantitatively in accordance with the requirements of Faraday's laws. If the c.d. is increased, however, a point is reached at which the anode potential rises suddenly, and there is a corresponding decrease of current; at the same time the nickel, or other, anode practically ceases to dissolve, although its appearance is unchanged. The metal is then said to be in the **passive state**.^{*} The general nature of the variations of potential is shown in Fig. 244. *AB* represents the change of anode poten-

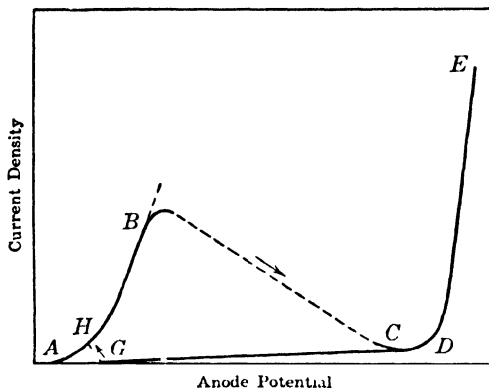


FIG. 244. Anodic passivity

tial with c.d., when the electrode is active and is dissolving quantitatively. At *B* the electrode becomes passive and the potential rises while the current decreases to *C*. If the c.d. is further increased, the potential becomes more anodic until a new process, generally oxygen evolution, commences at *D*, after which considerable current can flow with little further polarization (*DE*). Along *CE* the anode has almost, but not entirely, ceased to dissolve. When the c.d. is decreased from *E*, the active condition is not regained at *C*; the current must be reduced to a much lower value, e.g., *G*, before the potential changes to *H*, on the curve *AB*, and quantitative dissolution recommences. If the current is switched off while the anode is passive, it returns slowly to the active condition on standing. By using the passive metal as a cathode in a cell, or by touching it with a less noble metal, which also has the effect of making the passive metal cathodic, or by scratching the surface, and so producing short-circuited cells, the active state is rapidly regained. The attainment of passivity depends on the nature of the electrolyte. In the iron-group metals it is favored by alkalinity and the presence of oxidizing anions, e.g., nitrate, chlorate, etc., but chloride ions inhibit the onset of passivity. Increase of temperature and agitation increase the c.d. at which the anode becomes passive. The ease of passivation increases in the order iron, cobalt, nickel.

^{*} The term "passivity" was introduced by C. F. Schonbein (1836) who discovered the phenomenon with an iron anode.

A chromium anode can be rendered passive, although its behavior is somewhat different from that of the metals described above. The anode dissolves first to form Cr^{++} ions, but when it becomes passive, as the c.d. is increased, the potential rises (anodically) to +1.1 volt, and the chromium goes into solution in the hexavalent state, forming CrO_4^{--} ions.* No oxygen is evolved, simply because the process $\text{Cr} \rightarrow \text{CrO}_4^{--}$, possibly through the intermediate formation of Cr^{+++++} , requires a less positive potential than the discharge of OH^- ions. If the electrolyte contains iodide ions, the discharge of these can occur at a still lower potential, viz., +0.6 volt, and the metal then does not dissolve in the passive state, but iodine is deposited. Chromium becomes passive more easily in alkaline solution and temperature has an inhibiting effect, just as with the iron-group metals, but chloride ions have relatively little influence. The metal can retain its passivity for some time, even in acid solution, but if made a cathode the active condition is rapidly regained.

Molybdenum and tungsten are rendered passive more easily in acid than in alkaline solutions, and oxidizing agents inhibit the passivity of tin; it is clear, therefore, that each metal requires its own specific conditions if it is to be made passive. Platinum and gold are nearly always passive, and hence they are regarded as unattackable; both metals can, however, be made active by the use of suitable electrolytes. A gold anode will dissolve in neutral or acid chloride solutions, and platinum is attacked in ammoniacal and cyanide solutions; in each case, however, complete passivity sets in as the c.d. is increased.

Chemical Passivity.—Passivity can be produced without the action of an electric current, as was first found by J. Keir (1790) and later, independently, by G. Wetzlar (1827). If iron is dipped into concentrated nitric acid, there may be an instantaneous reaction, but the metal does not continue to dissolve, and although unchanged in appearance it no longer displaces silver from solution; the iron has thus become passive. The passivity can be removed by scratching the surface or by touching it, under the surface of an electrolyte, with a less noble metal, e.g., active iron. Chemical passivity can be induced by nitric acid and other oxidizing agents not only in iron, but also in nickel, cobalt and chromium. By heating in air these metals can be rendered passive, and a semi-passive state, in which the electrode potential is intermediate between active and passive values, can be induced by mere exposure to air.† The resemblances between a metal rendered passive by chemical and electrochemical methods is very marked, and there is little doubt that the fundamental cause of passivity is the same in each case.

* As with cathodic potentials (see footnote, p. 1023) the convention as to the sign of anode potentials in electrolysis is that for equilibrium reduction potentials. Anodic process, viz., solution of metals and discharge of anions, take place in the order of increasing anodic (positive) potentials.

† Stainless steel and chromium plate provide striking illustrations of passivity resulting from exposure to air.

Theories of Passivity.—Many theories have been proposed to explain the varied phenomena of passivity, and some modern views are a modification of those originally put forward by M. Faraday (1836); he suggested that the surface of a passive metal was oxidized, although it was realized that the layer of oxide might be so thin as to be invisible. This theory held its ground for many years, until it was shown that the surface oxide would need to have properties not usually associated with metallic oxides. Further, the supposed identity in light-reflecting powers of a passive anode and a cathode suggested that an oxide film could not be present on the former. In recent years, however, it has been shown that a passive electrode does not reflect plane polarized light to the same extent as an active electrode; hence a film of some kind is probably present on the surface. Further, photoelectric emission from passive iron is less than from the active metal. It is significant, too, that the oxides of iron, cobalt, etc. are less soluble in alkaline than in acid solution and passivity sets in more readily in the former. On the other hand, the oxides of molybdenum and tungsten are more soluble in an alkaline solution, and these metals can be made passive more easily in the presence of acid. By gentle anodic action, or by means of iodine, U. R. Evans (1927) succeeded in dissolving away the inner portion of passive iron, leaving a very thin, transparent and almost colorless film of oxide. The film is quite invisible when attached to the metal, but can be seen when separated from it. Although the existence of a thin layer of oxide on the surface of a passive metal may be admitted, it must be realized that it is still necessary to explain why dissolution ceases and how the oxide film is formed.

The work of W. J. Muller and his collaborators (1924 *et seq.*) has thrown much light on these problems. Using horizontal anodes, protected by means of a glass hood so that the product formed did not diffuse away, it was found that iron, nickel and chromium could be rendered passive by the use of smaller c.d.'s than had been previously regarded as effective. The time required for the onset of passivity is, however, relatively large, e.g., in *N* sulfuric acid an iron anode becomes passive in 12.4 sec. at a c.d. of 289 milliamp. per sq. cm., but if the c.d. is reduced to 15 milliamp. the time required is about 848 sec. If the solution is previously saturated with the corresponding salt of the anode metal, passivity can be brought about much more readily. These results suggest that a substance accumulates in the vicinity of the anode until the solution becomes saturated; then the solid separates out and deposits on the electrode. In several instances this deposit has been removed and identified as a known salt. A large part of the anode thus becomes shielded and the effective c.d. on the clear portions is much increased, and this eventually leads to the onset of passivity. It has been observed that immediately before passivation a visible film appears on the anode, but it is thrown off, or dissolves, when the electrode actually becomes passive and oxygen is evolved; the electrode then appears quite clean.

The mechanism proposed by E. S. Hedges (1928) and U. R. Evans (1929-30) to account for the onset and maintenance of passivity takes these facts into consideration. The liquid in the vicinity of an anode must sooner or later become saturated with a normal or basic salt of the metal, and the solid will deposit on the electrode surface. The film so formed is probably discontinuous and still allows the passage of metal ions into solution under the influence of the current. The result is that the deposit continues to grow in thickness and

covers most of the electrode; this is presumably the film observed, and sometimes identified, by Müller and others. The effective c.d. at the uncovered portions of the anode will be very large, and the potential rises until hydroxyl ion discharge can occur; the discharged ions react with the metal and produce an adherent *invisible* film beneath the visible, thick, but discontinuous, layer. The oxide film spreads rapidly across the electrode, and when it is covered, oxygen evolution commences; the bubbles of gas cause the first film to be thrown off leaving the surface apparently clean. The discontinuous film has played its part in reducing the effective area, and is not necessary for the maintenance of passivity. Since the invisible oxide film, which is presumably the one isolated after dissolving out the inner portion of passive iron, is adherent and continuous it cannot increase in thickness, and so remains invisible. The oxide is evidently formed in a very compact condition in which state it is not readily attacked by acids. Chloride ions are apparently able to penetrate certain films, e.g., ferric oxide, and render them porous and hence ineffective; the presence of chlorides thus retards the onset of passivity. Any factor tending to dissolve or reduce the adherent oxide film, such as making the passive metal a cathode, destroys the passive state. The important point to note is that whereas earlier workers were under the impression that a relatively thick film was necessary to ensure passivity, it is now realized that this cannot be so; for a thick film to form the metal must be able to continue to dissolve, and hence it is not passive. A coherent and adherent layer is essential if the metal is to be really passive, and such a film does not permit the passage of ions, and so it cannot grow in thickness.

Chemical passivity can be brought into line with the views expressed above. If iron is placed in nitric acid there will be rapid attack at first leading to a high local concentration of ferric ions and a depletion of hydrogen ions, with the result that a thin adherent film of ferric oxide, or a basic salt, is formed which protects the metal from further attack. The passivity produced in certain metals, particularly the chromium-iron alloys, e.g., stainless steel, on exposure to air is probably also due to a similar oxide film. It must not be imagined that chemical passivity is restricted to chromium and the iron group, and similar metals, for at -11° c. cobalt and copper can be rendered passive, and even zinc and magnesium show signs of passivity, in concentrated nitric acid. The decreased solubility of oxide or basic salt at the low temperature enables an adherent layer to be formed which cannot be obtained at ordinary temperatures. It appears, therefore, that passivity is a property exhibited by most metals under different conditions.

Mechanical Passivity.—In certain cases the dissolution of an anode is prevented by a visible film, e.g., lead dioxide on a lead anode in dilute sulfuric acid; this effect has been called "mechanical passivity," but it is probably not fundamentally different from electrochemical passivity. Lead first dissolves in the sulfuric acid solution in the form of plumbous ions, but a layer of insoluble lead sulfate is soon formed. The c.d. is increased so much that the dissolution of lead cannot occur sufficiently rapidly to satisfy the requirements of the current, and so the potential rises to that at which the process $\text{Pb} \rightarrow \text{Pb}^{++++} + 4e$ occurs. Unless the sulfuric acid is very concentrated, this results in the formation of a film of lead dioxide; it is, however, not entirely adherent and impermeable, and

so it increases in thickness and becomes visible. Such an oxide layer is not completely protective and attack of the anode continues to some extent.⁴²

Oxygen Evolution and Overvoltage.—After an electrode has become passive, oxygen evolution generally commences on increasing the c.d., unless another process can occur at a lower anodic potential, e.g., with chromium. The potential at which oxygen evolution commences is invariably greater (anodically) than the reversible value; there is consequently an overvoltage for oxygen evolution. The bubble overvoltages for a number of metals in *N* potassium hydroxide are given in Table 171.

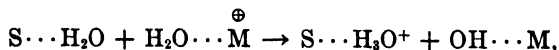
TABLE 171. OXYGEN OVERVOLTAGES IN *N* POTASSIUM HYDROXIDE

Nickel	0.06 volt	Silver	0.41 volt
Cobalt	0.14	Cadmium	0.43
Platinized Platinum	0.25	Palladium	0.43
Iron	0.25	Platinum (smooth)	0.45
Lead	0.31	Gold	0.53

It has been claimed (F. P. Bowden, 1929) that in a solution of dilute sulfuric acid saturated with oxygen, the overvoltage at a platinum anode varies with c.d. in the same manner as does hydrogen overvoltage [equation (270)], and that the constant *b* has the same value. If this is correct it implies a resemblance between oxygen and hydrogen overvoltages which had not been previously realized. It is important to point out that oxygen overvoltages are difficult to study because the potential of the anode does not remain steady. A further difficulty is that the surface usually becomes covered with an oxide film, of varying thickness and composition, during the course of electrolysis.

Until relatively recent times it was generally accepted that oxygen evolution at an anode occurred only through the intermediate formation of unstable "higher" oxides, e.g., Cu₂O₃ and CuO₂ on copper, Ni₂O₃ and NiO₂ on nickel, PtO₃ and PtO₄ on platinum, etc., and that these oxides were responsible for anodic overvoltage. It seems established from a study of the discharge potentials of various anodes, after the polarizing current is cut off, that such oxides are frequently present on the surface of an anode, but it is not certain that they are the cause of overvoltage. It is possible that at relatively high c.d.'s the measured anode potential is actually that of an unstable oxide, formed by the combination of the electrode material with active oxygen, but there still remains to be accounted for the fact that oxygen evolution does not commence at or near the reversible potential.

The similarity between hydrogen and oxygen overvoltages mentioned above led H. Eyring, S. Glasstone and K. J. Laidler (1939) to propose a mechanism for the liberation of oxygen at an anode which is quite analogous to that suggested for hydrogen evolution (p. 1022). Anodic overvoltage is attributed to the slowness of the transfer of a proton from a water molecule attached to the anode surface (M) to one associated with the electrolyte (S); thus,



where \oplus represents the removal of an electron from the metal. The OH

radicals left on the electrode by the discharge process react in pairs with the ultimate formation of oxygen gas. The connection between oxygen and hydrogen overvoltages is at once apparent. Since metals which form strong $M-H$ bonds will not permit the ready transfer of protons, such materials should have high oxygen overvoltages; this effect is opposite to that for cathodic overvoltage. The exact parallelism is confused by the formation of oxides on the anode, but it may be significant that platinum, palladium and gold have the lowest hydrogen overvoltages and the highest oxygen overvoltages.

It is of interest to call attention to a consequence of the theories presented here and on p. 1022. The essential process at both anode and cathode is the same, namely the transfer of a proton, and it is independent of the hydrogen ion concentration of the electrolyte. Since H is left on the cathode and OH on the anode, and these eventually give H_2 and O_2 , the ultimate reaction is, of course, the decomposition of water. The constancy of the decomposition voltage observed by LeBlanc (p. 1015) is thus to be expected.

Dissolution and Corrosion of Metals.—Any metal with an *oxidation* potential more positive than that of hydrogen in a given electrolyte should displace this element from solution; thus lead ($E^0 = +0.13$) should dissolve readily in hydrochloric acid with the evolution of hydrogen ($E^0 = \pm 0.0$), and zinc ($E^0 = +0.76$) should displace the latter from a neutral solution, e.g., sodium sulfate, in which the reversible hydrogen potential is about $+0.41$ volt. In actual practice neither of these reactions occurs, because of the hydrogen overvoltage which applies just as for evolution of the gas during electrolysis; the overvoltages of lead and zinc are 0.64 and 0.7 , respectively, and hydrogen evolution can only commence at -0.64 volt in hydrochloric acid for lead, and -1.11 volt for zinc in a neutral solution. By increasing the hydrogen ion concentration, the potential for gas evolution is made less negative, and so the tendency for the metal to dissolve is increased. Thus in a $2\ N$ solution of a strong acid the reversible hydrogen potential is about ± 0.0 , and so gas should be evolved at a zinc electrode at -0.7 volt; hence zinc would be expected to dissolve slowly in dilute acid. Actually impure zinc dissolves readily partly because its overvoltage is less than 0.7 , and partly for other reasons; pure zinc, however, in agreement with theory, is only attacked slowly by dilute acids.

If pure zinc is connected with a piece of copper and both are dipped into dilute acid, so as to form a short-circuited simple voltaic cell, the rate of dissolution of the zinc is markedly increased, but the hydrogen appears at the copper electrode. The less noble metal, zinc, dissolves to form zinc ions at one electrode (anode), and hydrogen ions are discharged at the copper (cathode); since the overvoltage at copper is only about 0.2 volt, hydrogen evolution occurs readily and the solution of zinc continues. In general, if the zinc is connected with a more noble metal having a low overvoltage it is attacked more readily. In order to expedite the solution of the zinc it is not necessary to set up a special cell, for the result may be achieved by bringing the metals in contact under the surface of an electrolyte; in this way the same voltaic cell is set up in a short-circuited form.

A similar effect can be obtained by adding a small quantity of copper sulfate to the solution in which the zinc is dissolving; copper is deposited, by replacement, on various parts of the zinc and a large number of local short-circuited elements are established. In these elements the zinc is the anode and dissolves, whereas the hydrogen is liberated on the copper deposit. Zinc may be taken as typical of a base metal with a high hydrogen overvoltage and copper of a more noble metal of low overvoltage, and the results obtained are of general applicability. Whenever a metal is in contact with, or has deposited on it, or included within it as a separate phase, a nobler metal having a small hydrogen overvoltage, the rate of solution of the base metal is increased. This mode of attack is sometimes called the "hydrogen evolution" type of corrosion, since gas is actually evolved at the more negative portions of the metal. It is important to remember that it can only occur if the oxidation potential of the base metal is more positive than that of a hydrogen electrode, allowing for the overvoltage at the more noble metal or inclusion. By reducing the acidity of the solution the potential of the latter is made more positive, and so it is often possible to increase the pH until the hydrogen evolution type of corrosion ceases.

Influence of Depolarizer.—If a depolarizer, which prevents the accumulation of hydrogen at the more noble electrode, is present the rate of attack of the base metal is increased. In these circumstances, since hydrogen gas is not evolved, the overvoltage of the noble metal is not so important as its nobility. Unless special precautions are taken to exclude air, oxygen is always available to act as a depolarizer; hence whenever two metals are in contact with a solution in air, the less noble metal will be attacked irrespective of the hydrogen overvoltage at either. For this reason the iron in imperfect "tin plate" is readily corroded in air, but in the absence of oxygen the high overvoltage at tin would prevent appreciable attack of the iron. Although corrosion is favored by a large difference of potential between positive and negative portions of a short-circuited voltaic cell, even the smallest of such differences is sufficient to stimulate corrosion in the presence of a depolarizer. In an apparently uniform piece of metal any portion which has been subjected to strain is generally less noble than an unstrained portion, and small crystals are less noble than large ones. Local galvanic elements of small E.M.F. can thus be set up in an electrolyte, and in the presence of a depolarizer corrosion of the more positive (baser) portions of the metal will occur. Short-circuited voltaic cells of this nature are probably always established when a single metal is attacked by an acid or other electrolyte.

Even noble metals can be corroded in the presence of a depolarizer; for example, copper is normally not attacked by dilute sulfuric acid, but if air is bubbled through the solution or hydrogen peroxide added to it, the copper dissolves. Normally the potential of the copper is too negative for it to be able to displace hydrogen from solution, although there must actually be a very minute pressure of the gas in equilibrium with copper in

an acid solution. In the presence of a depolarizer this hydrogen is oxidized, and so the copper continues to dissolve in the effort to attain equilibrium. In the language of thermodynamics, the hydrogen evolution reaction is replaced by one involving a larger decrease of free energy. The vigorous action of nitric acid on copper may be attributed to its depolarizing action on hydrogen; as is to be expected, reduction products of nitric acid are formed at the same time.

Differential Oxygenation.—An important type of corrosion results when the distribution of depolarizer, generally oxygen, is not uniform over the surface of the metal; this has been called corrosion due to "differential oxygenation." If oxygen has access to certain portions of a metal, depolarization of hydrogen will occur more readily there, and the potential will be more negative than on the other parts; the latter portions will, therefore, tend to dissolve because an E.M.F. has been set up. It seems strange, at first sight, that the parts of the metal which dissolve are those where the oxygen has not access, whereas the more highly oxygenated areas do not dissolve. This is because depolarization occurs at the latter parts, and hydrogen ions will be discharged there to replace the atoms removed by oxidation. Differential oxygenation accounts for corrosion at the bottom of "pits" in the surface of a metal, and for the preferential rusting of iron under those areas already covered by wet rust; in each case the corroded portion is that to which the air has limited access. The formation of oxide films, and the consequent passivation of the metal freely exposed to air, tend to make these portions more noble; the attack on the other parts is thus favored.⁴³

Electrolytic Oxidation and Reduction.—Since hydrogen tends to be evolved at a cathode and oxygen at an anode, it is evident that reduction and oxidation reactions, in the restricted sense of these terms, may be brought about at cathode and anode, respectively. In general, oxidation and reduction can be regarded as the removal and addition of electrons, and when a current from an external battery is applied to a pair of electrodes so as to make them anode and cathode, the former can act as a continuous remover of electrons, and hence bring about oxidation, whereas the latter will favor reduction since it is able to supply electrons. Electrolytic oxidation and reduction processes, as distinct from ionic discharge reactions, are thus very common and they are of many types; from the theoretical standpoint it is best to consider them under the headings of reversible and irreversible reactions.

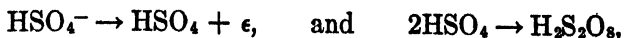
Reversible Oxidation-Reduction Reactions.—The principle involved is identical with that concerned in the reversible deposition or solution of metals. If the potential of a cathode is made just more negative than the reversible oxidation-reduction potential of the solution in which it is placed, reduction will commence; when the anodic potential exceeds the reversible value oxidation occurs. For example, if two platinum electrodes are placed in an acidified solution of equimolar amounts of a

ferrous and a ferric salt ($E^0 = -0.77$), and a gradually increasing E.M.F. applied, oxidation of ferrous to ferric ions will commence at the anode when its potential is just greater than 0.77 volt; at the cathode reduction will occur when the potential is just less than 0.77 volt. For small currents there will be little polarization, so that the potentials are close to the reversible values, but at high c.d.'s concentration polarization, due to an increase in ratio of ferric to ferrous ions at the anode and a decrease near the cathode, will be evident. At c.d.'s such that the maximum rate of oxidation or reduction is insufficient to satisfy the requirements of the current, the evolution of chlorine or oxygen will occur at the anode, or of hydrogen at the cathode. In some cases the oxidation-reduction potential may have a value that permits another process to take place simultaneously even at low c.d. The reversible potential in an acid titanous-titanic ion solution is about -0.05 volt, and at a platinized platinum electrode hydrogen evolution commences at ± 0.0 volt, approximately; hence at such a cathode reduction of titanous ions will be accompanied by considerable hydrogen evolution. By the use of a cathode material having an appreciable hydrogen overvoltage, however, evolution of the gas will not occur until a more negative potential, and the electrolytic reduction process can be carried out with 100 per cent efficiency. The cathodic reduction of stannous to stannous ions in hydrochloric acid solution is accompanied by liberation of hydrogen and also by the deposition of metallic tin, for all three processes take place in the vicinity of a deposition potential of -0.2 volt.

It should be noted that many reactions which might be expected to occur reversibly show appreciable polarization; instances are the reduction of V^V to V^{IV} and of V^{IV} to V^{III} , and also of Mo^{VI} to Mo^V and of Mo^V to Mo^{III} . In each of these cases there is marked irreversibility at a smooth platinum cathode, although it is much less with platinized platinum (F. Foerster, 1930). The abnormal behavior is probably due to the formation of insoluble oxides on the cathode; these have the effect of blocking most of the surface and thus giving a high effective c.d. Considerable polarization accompanies the reduction of chromate to chromic ions, and oxides are here also formed on the cathode. The reverse reaction, i.e., the oxidation of chromic to chromate ions, also appears to involve polarization effects.

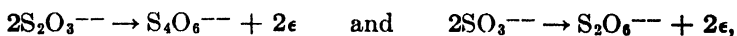
Polymerization of Anions.—The type of oxidation, of which the formation of persulfuric acid is one of the best known illustrations, involving the union of two anion residues, was at one time thought to be reversible. It is very doubtful, however, whether any of the known cases can be regarded as being reversible in the thermodynamic sense. As these processes have certain interesting properties in common they may be considered separately from nonreversible reactions. The electrolysis of relatively concentrated sulfuric acid, or of solutions of ammonium, potassium and sodium sulfates, gives persulfuric acid, or one of its salts, at the anode. It was originally assumed that the formation of persulfate was to

be attributed to the polymerization of HSO_4 radicals formed by the discharge of bisulfate ions; thus,



but this theory is not now universally accepted. It is true that there is a parallelism between the HSO_4^- concentration and the yield of persulfuric acid obtained in the electrolysis of sulfuric acid solutions, but even higher efficiencies are given by ammonium sulfate solutions, in which there is very little HSO_4^- . There is a possibility that persulfates may result from chemical oxidation of SO_4^{--} or HSO_4^- ions by active oxygen, or other oxidizing agent, e.g., hydrogen peroxide or hydroxyl radicals, formed at the anode.

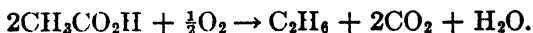
The anodic oxidation of thiosulfate to tetrathionate and of sulfite to dithionate, viz.,



may be formulated as electrical processes involving the direct discharge of $\text{S}_2\text{O}_3^{--}$ or SO_3^{--} ions, but there is a decided opinion in favor of a purely chemical mechanism. It has been suggested (S. Glasstone and A. Hickling, 1932-33) that hydroxyl radicals, formed at the anode by the discharge of hydroxyl ions, combine to form hydrogen peroxide, and that the hydrogen peroxide is the effective anodic oxidizing agent in many instances. It is also possible that the hydroxyl radicals may themselves be active in this respect. It is a striking fact, among many others in support of this theory, that in the presence of very small amounts of substances able to catalyze the decomposition of hydrogen peroxide,* the electrolytic formation of tetrathionate and dithionate can be almost entirely inhibited.

The conversion of carbonate to percarbonate and of phosphate to perphosphate are further examples of electrolytic polymerization of anions taking place at an anode. There is strong evidence in favor of the view that the effective oxidant is in each case a chemical oxidizing agent produced by the current.

The Kolbe and Allied Reactions.—The phenomena observed in the electrolysis of acetates have a number of characteristics similar to those found in the reactions just considered. Electrolysis of an aqueous solution of an acetate, alone or mixed with acetic acid, gives ethane and carbon dioxide, with a high efficiency, at the anode; the net reaction is



The process was discovered by H. Kolbe (1849) and is known by his name; it is a general reaction for most aliphatic, but not for aromatic, acids. The Kolbe synthesis occurs readily at a smooth platinum anode, but not at platinized platinum, palladium, gold or nickel. The presence

* Catalysts for hydrogen peroxide would presumably also favor the conversion of hydroxyl radicals into molecular oxygen and water.

of catalysts for the decomposition of hydrogen peroxide, or of appreciable amounts of other electrolytes, inhibits the formation of ethane. In each case, however, the current oxidizes the acetate, or acetic acid, to methyl alcohol; this is known as the Hofer-Moest reaction, for the possibility of obtaining methyl alcohol by the electrolysis of an acetate in the presence of other anions was first shown by H. Hofer and M. Moest (1902). The Kolbe reaction takes place in nonaqueous solutions, but the mechanism is evidently different from that in water; the nature of the electrode has now little influence and hydrogen peroxide catalysts do not affect the oxidation efficiency.

To explain the anodic oxidation of acetate ions, A. Crum Brown and J. Walker (1891) suggested that the ions were discharged directly at the anode and that the resulting radicals reacted; thus,



The main evidence for this view is that the evolution of ethane does not commence until a definite potential is attained, presumably that required for the discharge of acetate ions, but it does not account for the influence of electrode material or of added substances in aqueous solution, nor for the replacement of the Kolbe by the Hofer-Moest reaction. According to F. Fichter (1918 *et seq.*) active anodic oxygen oxidizes acetate ions or acetic acid to acetyl peroxide (CH_3CO_2)₂, which then decomposes to give ethane and carbon dioxide. There are many points in favor of this view, but it still leaves many observations unexplained. The theory that the effective oxidizing agent is the hydroxyl radical, or hydrogen peroxide formed by the combination of these radicals in pairs, appears to provide a means of combining the essential points of the two previous theories; it is also able to account for the influence of anode material, added salts, etc., on the Kolbe reaction. Under normal conditions the hydrogen peroxide (or hydroxyl) presumably reacts with acetate ions to form acetate radicals, which then combine, possibly with the intermediate formation of acetyl peroxide, to yield ethane and carbon dioxide. A small proportion of the current is also devoted to the direct discharge of acetate ions, which is followed by the reaction just mentioned. Catalysts for hydrogen peroxide decomposition, or electrode materials having the same effect, will consequently suppress the Kolbe reaction. The presence of added anions provides an alternative means for the transport of current, and so the concentration of acetate ions near the anode is kept low and the formation of ethane is inhibited. It is probable that in both these cases peracetic acid is produced which decomposes into methyl alcohol and carbon dioxide. In nonaqueous solutions hydrogen peroxide or hydroxyl radicals cannot be formed, since no hydroxyl ions are available, and hence direct discharge of acetate ions is the only possible process; the radicals probably combine to give acetyl peroxide which subsequently decomposes. It is readily understood, therefore, why the nature of the anode material exerts almost no influence on the reaction.

Nonreversible Oxidation and Reduction.*—Nonreversible electrolytic reactions occur most frequently with nonionized substances, e.g., organic compounds; the reduction of nitrobenzene to aniline, and oxidation of an alcohol to a carboxylic acid, are examples of reactions which may be carried out at a cathode and anode, respectively. A number of inorganic reactions, e.g., reduction of nitric acid and nitrates to hydroxylamine and ammonia, are also probably nonreversible. Although the electrolytic oxidation and reduction of organic compounds have been the subject of much study, the exact mechanism of the reactions involved is still uncertain. Reduction processes are perhaps better understood as fewer by-products are involved; cathode reactions will, therefore, be considered first. In general a nonreversible reaction may be regarded as a reaction between the reducible substance and atomic hydrogen; thus,



but it is possible that in some cases direct transfer of electrons from the cathode may be the primary process, which is followed by the addition of hydrogen ions



The velocity of the reduction process will depend, among other factors, on the concentrations of the reducible substance (depolarizer) and of atomic hydrogen, or its equivalent, and the latter will depend on the cathode potential; these variables will also determine the possibility of forming intermediate reduction products, e.g., ROH and RH.

The work of F. Haber (1898) has shown that by varying the cathode potential it is possible to obtain different reduction products with the same electrode. The less cathodic the potential the smaller is the concentration of active hydrogen at the electrode and the less efficient the reduction process. For example, at a c.d. of 0.001 amp. per sq. cm. the potential of a smooth platinum cathode in an alkaline alcoholic solution of nitrobenzene is -0.6 volt, and the chief reduction product is azoxybenzene; if the c.d. is increased to 0.0035, however, the cathodic potential rises to -1.0 volt and hydrazobenzene is formed. Overvoltage may also be expected to have an influence on electrolytic reduction. No matter what is its actual cause, the fact that a higher potential is attainable at a high overvoltage metal must mean a greater activity of the hydrogen, and hence a greater speed and efficiency of reduction. The hydrogen evolved from a low overvoltage cathode should have little more activity than the ordinary gas at atmospheric pressure. The results of J. Tafel (1900 *et seq.*) and others are in agreement with this expectation. When organic substances are difficult to reduce, e.g., carboxyl compounds and pyridine, the best results are obtained with lead, mercury and zinc cathodes. In the electrolytic reduction of nitro-compounds these metals generally yield

* The essential difference between a reversible and a nonreversible process is that the latter involves a slow stage, i.e., one with a high energy barrier, in one direction.

amines, but at nickel, carbon or platinum, intermediate products, e.g., phenyl hydroxylamine and benzidine, are usually obtained. Overvoltage is, however, not always the dominant factor in the behavior of an electrode, for in certain instances catalytic influences are important; nitrobenzene, for example, is easily reduced to aniline, and nitric acid to ammonia, at a copper cathode, in spite of its low overvoltage. Increase of temperature lowers overvoltage but increases the reaction rate; if the process is one requiring a high potential then the efficiency is generally decreased by raising the temperature. The concentration of the depolarizer does not usually influence the reaction products, but the reduction can be carried out with higher efficiency; increase of temperature often operates in the same direction. Cathodic reductions can be facilitated by the presence of catalysts which act as "hydrogen carriers"; instances are salts of titanium, vanadium and iron. The ion, e.g., Ti^{+++} , is first reduced at the cathode, e.g., to Ti^{++} , which is the active reducing agent; in the process it is re-oxidized to Ti^{+++} , and is again reduced at the cathode and so on.

Electrolytic Oxidation.—In general the factors influencing electrolytic reduction have an analogous effect in nonreversible oxidations. There is a rough parallelism between oxidation efficiency and anodic overvoltage in certain reactions, e.g., oxidation of formate ions, but there are others, e.g., oxidation of oxalate and nitrite, where the catalytic influence of the anode material is the dominant factor. As a rule platinum and lead, effectively PbO_2 , anodes are the most efficient and nickel the least effective. Greatly improved results are often obtained by increasing the electrode area, for at platinized platinum both methyl alcohol and oxalic acid are oxidized more readily than at a smooth platinum anode. It is possible that organic electrolytic oxidations may be divided into two broad classes, viz., those in which hydrogen peroxide (or hydroxyl) is the actual oxidizing agent, and those in which active oxygen is effective. The former will occur most readily at metals which are poor catalysts for hydrogen peroxide decomposition, e.g., smooth platinum, whereas the others may be favored by anodes which decompose the peroxide, e.g. lead dioxide and platinized platinum. Variation of the potential of a given anode sometimes affects the nature of the products, a low potential giving a less highly oxidized material; e.g., ethyl alcohol gives mainly acetaldehyde at low potential and acetic acid at high potential, with a smooth platinum anode in acid solution. Increase of temperature usually brings about an increase in efficiency of nonreversible anodic oxidations and so also does an increase in the concentration of depolarizer. Finally, cerium, vanadium, manganese and chromium compounds can often act, as "oxygen carriers" in a manner similar to the action of the hydrogen carriers already described.⁴⁴

REFERENCES

- Britton, H. T. S. "Hydrogen Ions."
- Clark, W. M. "The Determination of Hydrogen Ions."
- Davies, C. W. "The Conductivity of Solutions."
- Dole, M. "Experimental and Theoretical Electrochemistry."
- Evans, U. R. "Metallic Corrosion, Passivity and Protection."
- Falkenhagen, H. "Electrolytes." (Trans. by R. P. Bell.)
- Glasstone, S. "The Electrochemistry of Solutions"; "Introduction to Electrochemistry."
- Gurney, R. W. "Ions in Solution."
- Harned, H. S. and Owen, B. B. "The Physical Chemistry of Electrolytic Solutions."
- Kolthoff, I. M. "Acid-Base Indicators." (Trans. by C. Rosenblum.)
- Latimer, W. M. "The Oxidation States of the Elements, etc."
- MacInnes, D. A. "The Principles of Electrochemistry."
- Lewis, G. N. and Randall, M. "Thermodynamics, etc."
- Taylor, H. S. "Treatise on Physical Chemistry." 2nd ed., Chap. XI (J. R. Partington); XII (H. S. Harned); XIII (N. H. Furman).
- Walden, P. "Salts, Acids and Bases, etc."
- Richards, *et al.*, *Z. phys. Chem.*, **41**, 302 (1902); **42**, 621 (1903); Datta and Dhar, *J. Am. Chem. Soc.*, **38**, 1156 (1916); Marshall, *Trans. Farad. Soc.*, **21**, 297 (1925); Matthews and Wark, *J. Phys. Chem.*, **35**, 2345 (1931).
 - Arrhenius, *Z. phys. Chem.*, **1**, 481 (1887); *Phil. Mag.*, **26**, 81 (1888).
 - For recent work see Jones, *et al.*, *J. Am. Chem. Soc.*, **50**, 1049 (1928); **51**, 2407 (1929); **55**, 1780 (1933); **57**, 272 (1935); Shedlovsky, *ibid.*, **52**, 1793 (1930); **54**, 1411 (1932); Luder, *ibid.*, **62**, 89 (1940).
 - Kohlrausch, *Ann. Phys.*, **6**, 1 (1879).
 - Walden, *Z. phys. Chem.*, **100**, 512 (1922); Fuoss and Kraus, *J. Am. Chem. Soc.*, **55**, 2387 (1933).
 - Walden, *Z. phys. Chem.*, **55**, 207, 246 (1906); Ulich, *Trans. Farad. Soc.*, **23**, 388 (1927); Hartley and Barak, *Z. phys. Chem.*, **165**, 273 (1933); Huggins, *J. Am. Chem. Soc.*, **53**, 3190 (1931).
 - Küster, *Z. anorg. Chem.*, **35**, 454 (1903); **42**, 225 (1904); Kolthoff, *ibid.*, **111**, 112 (series of papers) (1920); Davies, *Trans. Farad. Soc.*, **29**, 429 (1933); Harned, *J. Am. Chem. Soc.*, **39**, 252 (1917); **40**, 1213 (1918); for reviews, see Glasstone, *Ann. Reports Chem. Soc.*, **30**, 294 (1933); Britton, "Conductometric Analysis."
 - Sherrill and Noyes, *J. Am. Chem. Soc.*, **48**, 1861 (1926); MacInnes, *ibid.*, p. 2068.
 - Debye and Hückel, *Phys. Z.*, **24**, 305 (1923); **25**, 145 (1924); *Trans. Farad. Soc.*, **23**, 334 (1927); Onsager, *Phys. Z.*, **27**, 388 (1926); **28**, 277 (1927); *Trans. Farad. Soc.*, **23**, 341 (1927).
 - For reviews, see Falkenhagen and Williams, *Chem. Rev.*, **6**, 317 (1929); Williams, *ibid.*, **8**, 303 (1931); MacInnes, *et al.*, **13**, 29 (1933); *J. Frank. Inst.*, **225**, 661 (1938); Hartley, *et al.*, *Ann. Rep. Chem. Soc.*, **27**, 326 (1930); Falkenhagen, *op. cit.*
 - Davies, *Trans. Farad. Soc.*, **23**, 351 (1927); Banks, *J. Chem. Soc.*, 3341 (1931).
 - Debye and Falkenhagen, *Phys. Z.*, **29**, 121, 401 (1928); Sack, *ibid.*, p. 627; Wien, *Ann. Phys.*, **83**, 327 (1927); **85**, 795 (1928).
 - Lodge, *Brit. Assoc. Rep.*, 389 (1886); Whetham, *Phil. Trans. Roy. Soc.*, **184A**, 337 (1893); **186A**, 507 (1895).
 - Nernst and Loeb, *Z. phys. Chem.*, **2**, 948 (1888); Hopfgartner, *ibid.*, **25**, 115 (1898); Jahn, *ibid.*, **37**, 673; Noyes, *ibid.*, **36**, 63 (1901); *J. Am. Chem. Soc.*, **23**, 148 (1901); Getman, *ibid.*, **36**, 1640 (1914); Jones, *et al.*, *ibid.*, **51**, 1073 (1929); **54**, 138 (1932); MacInnes and Dole, *ibid.*, **53**, 1357 (1931).
 - MacInnes, *et al.*, *ibid.*, **45**, 2246 (1923); **46**, 1398 (1924); **47**, 994, 1009 (1925); **48**, 1909 (1926); Cady and Longworth, *ibid.*, **51**, 1656 (1929); Longworth, *ibid.*, **57**, 1698 (1935); MacInnes and Longworth, *Chem. Rev.*, **11**, 181 (1932).
 - Washburn, *J. Am. Chem. Soc.*, **31**, 322 (1909); **37**, 694 (1915); Taylor, *et al.*, *J. Chem. Soc.*, 2497 (1932); 902 (1937).

17. Sand, *ibid.*, **91**, 380 (1907); *J. Soc. Chem. Ind.*, **30**, 37 (1911); Hughes, *J. Am. Chem. Soc.*, **44**, 2860 (1922); Elder, *ibid.*, **51**, 3266 (1929); Dole, *ibid.*, **53**, 620 (1931); Morton, *J. Chem. Soc.*, 2977, 2983 (1931); Garman and Drosz, *Ind. Eng. Chem. (Anal.)*, **11**, 398 (1939).
18. MacInnes and Beattie, *J. Am. Chem. Soc.*, **42**, 1117 (1920); Jones and Dole, *ibid.*, **51**, 1073 (1929); Jones and Bradshaw, *ibid.*, **54**, 138 (1932); Harned and Dreby, *ibid.*, **61**, 3113 (1939).
19. Meyer, *Z. phys. Chem.*, **7**, 447 (1891); Lewis and Rupert, *J. Am. Chem. Soc.*, **33**, 299 (1911); Richards, *et al.*, *ibid.*, **41**, 1732 (1919); **44**, 601 (1922); Loomis, *et al.*, *ibid.*, **38**, 2391 (1916); *J. Phys. Chem.*, **21**, 334 (1917); MacInnes, *et al.*, *J. Am. Chem. Soc.*, **44**, 1021 (1922); **46**, 1437 (1924); Romann and Chang, *Bull. Soc. chim.*, **51**, 932 (1932).
20. Planck, *Ann. Phys.*, **40**, 561 (1890); Henderson, *Z. phys. Chem.*, **59**, 118 (1907); **63**, 325 (1908); Guggenheim, *J. Am. Chem. Soc.*, **52**, 1315 (1930); Lewis and Sargent, *ibid.*, **31**, 363 (1909); MacInnes and Yeh, *ibid.*, **43**, 2563 (1921); Lamb and Larson, *ibid.*, **42**, 229 (1920); Lakhani, *J. Chem. Soc.*, 179 (1932).
21. Nernst, *Z. phys. Chem.*, **4**, 129 (1889); Gurney, *Proc. Roy. Soc.*, **136A**, 378 (1932); *op. cit.*; Butler, "Electrocapillarity," Chap. III; Glasstone, Laidler and Eyring, "The Theory of Rate Processes," Chap. X.
22. Bancroft, *Z. phys. Chem.*, **10**, 837 (1892); Peters, *ibid.*, **26**, 193 (1898); Haber and Russ, *ibid.*, **47**, 257 (1904); for reviews, see Clark, *Chem. Rev.*, **2**, 127 (1925); *J. App. Phys.*, **9**, 97 (1938); Michaelis, *Chem. Rev.*, **16**, 243 (1935); "Oxydations-Reductions Potentiale."
23. Ostwald, *Z. phys. Chem.*, **2**, 36, 270 (1888); **3**, 170, 241, 369 (1889); Bredig, *ibid.*, **13**, 289 (1894).
24. Debye and Hückel, *Phys. Z.*, **24**, 185, 334 (1923); **25**, 97 (1924); *Rec. trav. chim.*, **42**, 597 (1923); Hückel, *Phys. Z.*, **26**, 93 (1925); for reviews, see Noyes, *J. Am. Chem. Soc.*, **46**, 1080 (1924); **48**, 621 (1926); Williams, *ref. (10)*; Falkenhagen, *op. cit.*
25. Gronwall, La Mer, *et al.*, *Phys. Z.*, **29**, 358 (1928); *J. Phys. Chem.*, **35**, 2345 (1931); Fuoss and Kraus, *J. Am. Chem. Soc.*, **55**, 1019, 2387 *etc.* (1933); for reviews, see *Chem. Rev.*, **17**, 227 (1935); *Trans. Am. Elec. Soc.*, **66**, 385 (1934); *J. Frank. Inst.*, **225**, 687 (1938).
26. Nernst, *Z. phys. Chem.*, **4**, 372 (1889); Noyes, *et al.*, *ibid.*, **6**, 241 (1890); *J. Am. Chem. Soc.*, **33**, 1643, 1650, 1671 (1911); Hill, *ibid.*, **32**, 1186 (1910); Harkins and Pearce, *ibid.*, **38**, 2679 (1916); Kendall and Andrews, *ibid.*, **43**, 1545 (1921).
27. Bodländer, *Z. phys. Chem.*, **39**, 597 (1902); *Z. anorg. Chem.*, **31**, 458 (1902); **39**, 197 (1904); Sherrill, *Z. phys. Chem.*, **43**, 705 (1903).
28. Lowry, *Chem. and Ind.*, **42**, 43 (1923); Brønsted, *Rec. trav. chim.*, **42**, 718 (1923); *J. Phys. Chem.*, **30**, 777 (1926); for reviews, see Brønsted, *Chem. Rev.*, **3**, 231 (1928); Hall, *ibid.*, **8**, 191 (1931); Bell, *Ann. Rep. Chem. Soc.*, **31**, 71 (1934).
29. Walker, *Z. phys. Chem.*, **4**, 319 (1889); *J. Chem. Soc.*, **67**, 576, (1895); **83**, 484 (1903); Bredig, *Z. phys. Chem.*, **13**, 213, 321 (1894); Farmer, *J. Chem. Soc.*, **79**, 863 (1901); **85**, 1713 (1904); Williams and Soper, *ibid.*, 2469 (1930); Guleczian and Miller, *J. Am. Chem. Soc.*, **54**, 3151 (1932).
30. Kohlrausch and Heydweiller, *Z. phys. Chem.*, **14**, 317 (1894); Arrhenius, *ibid.*, **11**, 805 (1893); Dawson, *J. Chem. Soc.*, 1146, 1290 (1927); Lewis, Brighton and Sebastian, *J. Am. Chem. Soc.*, **39**, 2245 (1917); Roberts, *ibid.*, **52**, 3877 (1930); Harned and Owen, *Chem. Rev.*, **25**, 31 (1939).
31. Clark, *op. cit.*; Britton, *op. cit.*; Dole, "The Glass Electrode."
32. Walker, *Proc. Roy. Soc.*, **73A**, 155; **74A**, 271 (1904); *Z. phys. Chem.*, **49**, 82 (1904); **51**, 706 (1905); Bredig, *ibid.*, **13**, 214, 323 (1894); *Z. Elek.*, **6**, 35 (1899); Adams, *J. Am. Chem. Soc.*, **38**, 1503 (1916); Bjerrum, *Z. phys. Chem.*, **104**, 147 (1923); Harris, *Biochem. J.*, **24**, 1080, 1086 (1930); Richardson, *Proc. Roy. Soc.*, **115B**, 121 (1934); Neuberger, *ibid.*, p. 180; Schmidt, "The Chemistry of Amino Acids and Proteins," Chap. XI (Hitchcock).

33. Acree, *Am. Chem. J.*, **38**, 19 (1907); Noyes, *J. Am. Chem. Soc.*, **32**, 815 (1910); Gillespie, *ibid.*, **42**, 742 (1920); Sørensen, *Biochem. Z.*, **21**, 131, 201 (1909); Clark and Lubs, *J. Biol. Chem.*, **2**, 1, 109, 191 (1917); Clark, *op. cit.*; Kolthoff, *op. cit.*
34. Sørensen, *loc. cit.*, ref. (32); Henderson, *J. Am. Chem. Soc.*, **30**, 954 (1908); Van Slyke, *J. Biol. Chem.*, **52**, 525 (1922); for details of preparation and use of buffer solutions, see Clark, *op. cit.*; Britton, *op. cit.*
35. For reviews of theory and methods of potentiometric titrations, see Kolthoff and Furman, "Potentiometric Titrations"; Furman, *Ind. Eng. Chem. (Anal.)*, **2**, 213 (1930); **14**, 367 (1942); *Trans. Elec. Soc.*, **76**, 45 (1939); Glasstone, *op. cit.*; *Ann. Rep. Chem. Soc.*, **30**, 283 (1933); Sutton, "Volumetric Analysis," Part V (Glasstone).
36. Lewis and Randall, *op. cit.*; La Mer and Cowperthwaite, *J. Am. Chem. Soc.*, **53**, 4333 (1931); **55**, 1004 (1933); Harned, *et al.*, *ibid.*, pp. 2179, 4838; **59**, 496, 1290 (1937); Latimer, *et al.*, *ibid.*, **48**, 2297, 2868 (1926); **60**, 1829 (1938); for summaries, see *J. Chem. Phys.*, **2**, 82 (1934); *Chem. Rev.*, **18**, 349 (1936).
37. LeBlanc, *Z. phys. Chem.*, **8**, 299 (1891); **12**, 333 (1893); Caspari, *ibid.*, **30**, 89 (1899); Tafel, *ibid.*, **50**, 641 (1905).
38. Glasstone, *J. Chem. Soc.*, 125, 2414, 2646 (1924); *Trans. Farad. Soc.*, **21**, 36 (1925); Tafel, ref. (37); Bowden and Rideal, *Proc. Roy. Soc.*, **120A**, 59 (1928); Hickling, *Trans. Farad. Soc.*, **36**, 364 (1940).
39. Tafel, *Z. phys. Chem.*, **34**, 200 (1900); **50**, 641, 713 (1905); Lewis and Jackson, *ibid.*, **56**, 193 (1906); Brunner, *ibid.*, p. 331; Bonhoeffer, *ibid.*, **113**, 199 (1924); Bowden and Rideal, ref. (38); Bowden, *Proc. Roy. Soc.*, **125A**, 446; **126A**, 107 (1929); Erdey-Grúz and Volmer, *Z. phys. Chem.*, **150**, 203 (1930); Gurney, *Proc. Roy. Soc.*, **134A**, 137 (1931); Hammett, *Trans. Farad. Soc.*, **29**, 770 (1933); Eyring, Glasstone and Laidler, *J. Chem. Phys.*, **7**, 1053 (1939); Kimball, Glasstone and Glassner, *ibid.*, **9**, 91 (1941); de Bèthune and Kimball, *ibid.*, **13**, 53 (1945); Hickling and Salt, *Trans. Farad. Soc.*, **38**, 474 (1942); for reviews, see Glasstone, *op. cit.*; Adam, "The Physics and Chemistry of Surfaces," 1938, Chap. VIII; Bowden and Agar, *Ann. Rep. Chem. Soc.*, **35**, 90 (1938); Wirtz, *Z. Elek.*, **44**, 303 (1938).
40. Blum and Hogaboom, "Electroplating and Electroforming," Chap. VII; Glasstone, *op. cit.*; General Discussion, *Trans. Elec. Soc.*, Special Volume, 1942.
41. Heyrovský, *Phil. Mag.*, **45**, 303 (1923); *Trans. Farad. Soc.*, **19**, 692 (1924); *Rec. trav. chim.*, **46**, 496 (1925); *Mikrochem.*, **12**, 25 (1932); for reviews, see Kemula, *Z. Elek.*, **37**, 779 (1931); von Stackelberg, *ibid.*, **45**, 466 (1939); Müller, *Chem. Rev.*, **24**, 95 (1939); Kolthoff and Lingane, *ibid.*, p. 1; "Polarography."
42. Müller, *Z. Elek.*, **30**, 401 (1924); **33**, 401 (1927); *Trans. Farad. Soc.*, **27**, 737 (1931); Hedges, *J. Chem. Soc.*, 2878 (1926); 969 (1928); Evans, *ibid.*, 1020 (1927); 109 (1929); 478 (1930); for references, see Heathcote, *J. Soc. Chem. Ind.*, **26**, 899 (1907); Gerding and Karssen, *Z. Elek.*, **31**, 135 (1925); Evans, *Nature*, **128**, 1062 (1931); *op. cit.*; Hedges, *Ann. Rep. Chem. Soc.*, **31**, 127 (1934); Uhlig, *Trans. Elec. Soc.*, **85** (1944).
43. Cushman, *Trans. Am. Elec. Soc.*, **12**, 403 (1907); Walker, Cederholm and Bent, *J. Am. Chem. Soc.*, **29**, 1256 (1907); Aston, *Trans. Am. Elec. Soc.*, **29**, 449 (1916); Evans, *J. Inst. Metals* (London), **30**, 61 (1923); *J. Soc. Chem. Ind.*, **43**, 129T, 315T (1924); **45**, 38T (1926); *Ind. Eng. Chem.*, **17**, 363 (1925); *op. cit.*
44. See Glasstone, *op. cit.*; Glasstone and Hickling, "Electrolytic Oxidation and Reduction"; *Chem. Rev.*, **25**, 407 (1939); Hickling, *et al.*, *J. Chem. Soc.*, 256 (1940); *ibid.*, 90 (1943).

CHAPTER XIII

CHEMICAL KINETICS

ORDER OF REACTION

Introduction.—All chemical reactions take place at a definite rate depending on the conditions, of which the most important are concentration of reactants, temperature, radiation and presence of a catalyst or inhibitor. Some reactions are so rapid that they appear to be instantaneous, e.g., neutralization of a strong acid by a strong base, whereas others are so slow at ordinary temperatures, e.g., the combination of hydrogen and oxygen, that no detectable change would be observed in the course of years. Between these two extremes are to be found many processes involving both inorganic and organic compounds, although mainly the latter, taking place with measurable velocities at temperatures that are easily accessible in the laboratory. It is the study of the rates at which such reactions occur, and the influence of the conditions on these rates, that is included in the subject of **chemical kinetics**. The earliest experiments in this connection were made by L. Wilhelmy (1850), to whose work reference was made on p. 816, and by A. V. Harcourt and W. Esson (1867), and although much work has been done in this field it is only during the present century that there has been some clarification of the complexities of reaction kinetics.

Order of Reaction.—Since the rate of a chemical change is proportional to the concentrations of the reacting substances (p. 817), it is evident that the velocity * of the process must fall off as the reaction proceeds, for the reactants are being continuously consumed. The change of reaction rate with time must be represented by a curve similar to that shown in Fig. 245, the reaction becoming slower and slower but never ceasing; theoretically, the curve should, in fact, be asymptotic to the time axis. It is obviously not a simple matter to define the general rate of a reaction, and so in practice the velocity is considered at a particular instant; in

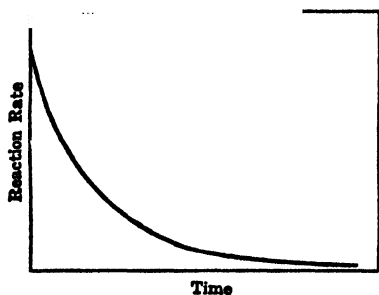
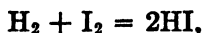


FIG. 245. Change of rate of reaction with time

this way valuable results can be obtained. In the study of chemical kinetics reactions are divided into classes determined either by molec-

* In chemical kinetics, the terms velocity, rate and speed are assumed to be synonymous and are employed indiscriminately.

ularity, that is, by the number of atoms or molecules taking part in each act leading to chemical reaction, or by the order of the reaction, that is, the number of atoms or molecules whose concentrations determine the velocity, or kinetics, of the process. Until quite recently these two terms were regarded as synonymous, but it now appears necessary to make some distinction, although the molecularity and order of a reaction are often identical. As will be seen shortly, the rate of thermal decomposition of nitrogen pentoxide is directly proportional to its concentration; the reaction is therefore said to be of the first order. It is also unimolecular, for there is good reason for believing that each elementary act of decomposition involves only one molecule. Similarly, the gas reaction between hydrogen and iodine, viz.,



is of the second order, since the rate of formation of hydrogen iodide is proportional to the concentrations of both hydrogen and iodine; the process is also bimolecular, each step involving the interaction of a molecule of hydrogen and one of iodine. On the other hand, although the process



is bimolecular, it is *not* found experimentally to be of the second order; the observed order is in fact 1.5, for reasons which will be explained later (p. 1060). Another illustration of the difference between the two classifications of reactions is supplied by the photochemical union of hydrogen and chlorine to form hydrogen chloride, when carried out over water saturated with the two reacting gases, as in the actinometer of R. Bunsen and H. E. Roscoe (1855); the reaction is very complicated (see p. 1173) although in a general sense it is apparently bimolecular, but it is kinetically of zero order, for the solution of the hydrogen chloride as fast as it is formed reduces the volume and keeps the concentration of the reacting gases constant. For the purpose of treating reaction rates mathematically it is most convenient to classify reactions according to their orders, to point out, where necessary, any difference between the order and the molecularity, and if possible to explain the discrepancy.

It is hardly necessary to emphasize the fact that chemical reactions are not always simple; complications as a result of reversible reactions, of simultaneous and side reactions, and of other causes, often arise. It will be convenient, however, to consider in the first place what have been called *isolated reactions*, which are free from any secondary effects; the influence of disturbing factors will be examined later. Further, for the present, only homogeneous reactions will be discussed, for it will be shown subsequently that special conditions arise in processes taking place on surfaces. It is the failure to distinguish between the orders of homogeneous and of heterogeneous reactions that was one of the causes of misleading results being obtained in the earlier studies of reaction

kinetics. Unless otherwise stated, constant temperature will be assumed in the treatment which follows.

Reactions of the First Order.—In a first order reaction the rate is directly proportional to the concentration of the reacting substance; the condition can be expressed mathematically in the form

$$-\frac{dc}{dt} = kc, \quad (1)$$

where c is the concentration of the reacting substance, for this equation states that the rate of disappearance of the reactant at any instant is proportional to its concentration at that instant. If a is the initial concentration of the reacting species, and x is the decrease after the lapse of time t , the concentration remaining will then be $a - x$; the reaction velocity can be obtained by substituting $a - x$ for c in (1); thus,

$$-\frac{d(a-x)}{dt} = k(a-x) \quad \text{or} \quad \frac{dx}{dt} = k(a-x). \quad (2)$$

The quantity dx/dt , which is a measure of the rate of decomposition, is called the reaction velocity, and the proportionality constant k^* is referred to as the **specific reaction rate**. (S. Arrhenius, 1887), or sometimes as the velocity constant or velocity coefficient. The reaction velocity at any instant is thus equal to the product of the specific rate, which is constant at a definite temperature for the given reaction, and the concentration of the reactant at that instant. In order to see how the velocity constant may be evaluated, it is necessary to integrate (2) in the form

$$\frac{dx}{a-x} = kdt, \quad (3)$$

remembering that when t is 0, i.e., at the beginning of the reaction, the amount changed x is 0, whereas after time t it is x ; the result is

$$k = \frac{1}{t} \ln \frac{a}{a-x} = \frac{2.303}{t} \log \frac{a}{a-x}. \quad (4)$$

This is known as the kinetic equation for a reaction of the first order, and for such a reaction the insertion of the values for a , and for $a - x$ corresponding to different times t during the course of the reaction, should give a constant value of k in (4). The same equation may be written (L. Wilhelmy, 1850) in the form

$$a - x = ae^{-kt} \quad \text{or} \quad c_t = c_0 e^{-kt}, \quad (5)$$

where c_0 is the initial concentration of reactant and c_t the value after time t . It is clear from this expression that the quantity of reacting material remaining falls off exponentially, and theoretically the reaction is com-

* This constant includes the factor \mathcal{F} for the medium, considered on p. 826.

plete, i.e., $c_t = 0$, only when t is infinite; in other words, the process is never absolutely complete. In practice, of course, most reactions taking place at a reasonable rate are virtually complete, within the limits of ordinary analytical detection, in the course of a short time.

Before proceeding to consider examples of reactions of the first order, it is of interest to examine (4) further. In the first place, it is seen that the quantity $a/(a - x)$ is a ratio of concentrations, and so it will be independent of the units used to express these concentrations provided, of course, the same units are used for both. This fact makes the evaluation of the first order velocity constant particularly simple, for the values of a and $a - x$ can be given in terms of moles per liter, or of ml. of a volumetric reagent, or of partial pressure of gas, or even as the degrees of rotation in a polarimeter. A further consequence is that the specific rate k is expressed in reciprocal time units; if the time t is given in seconds, as is frequently the case, k is in terms of sec.^{-1} .

Another important conclusion is reached when the time taken to complete a definite fraction of the reaction is calculated; for example, let $t_{0.5}$ be the time for the concentration to be reduced to half its initial value, i.e., x is then equal to $a/2$. Inserting this value for x in (4) it follows that

$$t_{0.5} = \frac{\ln 2}{k}, \quad (6)$$

so that the time is independent of the initial concentration; the same result is obtained for the time to reduce the concentration by any other given fraction. This is one of the most striking facts about reactions of the first order: the time taken to reduce the concentration of reactant from 1 mole per liter to 0.5 mole per liter is exactly the same as that required to decrease it from 0.001 to 0.0005 mole per liter, and so on. The decay of a radioactive element may be regarded as a first order process; it shows the characteristic property of a definite period of half-change which is independent of the initial amount of material (p. 129).

To test whether a particular reaction is of the first order or not, it is necessary to see if the experimental results satisfy (1). Sometimes the initial concentration is not known definitely because the exact moment when the reaction commences is uncertain. The value of a itself may then be eliminated by taking two different values of $a - x_1$ and $a - x_2$ at time intervals of t_1 and t_2 respectively; then,

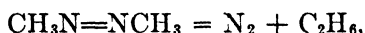
$$k = \frac{2.303}{t_2 - t_1} \log \frac{a - x_1}{a - x_2} = \frac{2.303}{\Delta t} \log \frac{c_1}{c_2}, \quad (7)$$

where c_1 and c_2 are the concentrations of the reactant at two instants differing by Δt in time. Instead of calculating k for various time intervals, a graphical method may be employed to cover the whole reaction period. By rearrangement (4) becomes

$$t = \frac{2.303}{k} \log a - \frac{2.303}{k} \log (a - x), \quad (8)$$

and since the first term on the right-hand side is constant, it follows that for a reaction of the first order the plot of t against $\log(a - x)$ should be linear; the slope of the line is equal to $-2.303/k$, thus enabling the specific reaction rate to be evaluated.

First Order Gas Reactions.—In 1924 only one homogeneous first order gas reaction, namely the thermal decomposition of nitrogen pentoxide vapor, was known (F. Daniels and E. H. Johnston, 1921); the decomposition of phosphine which until that time was believed to be homogeneous and unimolecular was shown to be heterogeneous (C. N. Hinshelwood and B. Topley, 1924). Shortly afterward, however, C. N. Hinshelwood (1926) showed that at temperatures of about 500° to 600° c. acetone and propaldehyde appeared to decompose in a unimolecular manner and that the reactions were homogeneous. In more recent times a considerable number of first order gas reactions have been discovered; mention may be made of the thermal decomposition of azomethane and of azoisopropane, giving nitrogen and a hydrocarbon (H. C. Ramsperger, 1927), e.g.,



of propyl-, isopropyl- and dimethyl-amines (H. A. Taylor, 1930), of nitrous oxide (M. Volmer, 1930), and of various ethers. Most of these reactions are, however, not really unimolecular, but are complex processes probably involving chain mechanisms (see p. 1111), although they appear to be kinetically of the first order. An entirely different type of process which has been found to be unimolecular is the isomerization of *d*-pinene to dipentene (D. F. Smith, 1927). In all these cases introduction of glass wool, or powdered silica, or other material having a large surface area, is said to have a relatively small effect on the reaction velocity, thus showing the reaction to be homogeneous. This test must always be applied to make certain that the process is not a heterogeneous one taking place on the surface of the vessel, for if it were, the presence of material of large surface area would increase the speed of the reaction. On the other hand, the rate of a chain reaction is often decreased by increasing the area of exposed surface (p. 1081).

The kinetics of gas reactions are most conveniently studied by observing the pressure of the gas after various intervals of time, provided the reaction is accompanied by a change in the number of molecules. The decomposition of nitrogen pentoxide may be represented ultimately by the equation



intermediate steps being ignored, so that the partial pressure of nitrogen pentoxide decomposed, i.e., x , should be twice the partial pressure of the oxygen formed at any instant. Unfortunately, the reaction is complicated by the dissociation of nitrogen tetroxide to the dioxide, which causes a still further increase of pressure; from a knowledge of the equilibrium constant of this reaction F. Daniels and E. H. Johnston (1921) were able

to make the necessary corrections and so to calculate the partial pressure of oxygen from the total measured pressure during the course of the reaction. In this manner the results quoted in Table 172 were obtained for k ,

TABLE 172. THERMAL DECOMPOSITION OF NITROGEN PENTOXIDE AT 45° C.

(Initial Pressure $a = 348.4$ mm.)					
t	x	k (sec. ⁻¹)	t	x	k (sec. ⁻¹)
20 min.	163.2 mm.	4.82×10^{-4}	80 min.	315.3 mm.	4.75×10^{-4}
40	243.0	4.78	100	329.8	4.52
60	289.8	4.85	160	345.6	4.67

determined by (4) with concentrations expressed as partial pressures. The approximate constancy of the k values indicates that the reaction is of the first order; if the data are inserted in the equations for reactions of higher order the results show a very marked drift.¹

Flow Method.—In addition to the method whereby the rate of a gas reaction is followed by increase of pressure, a "flow method" has been employed. The gas under investigation is passed at a known rate through a decomposition chamber, consisting of a wide tube of known volume kept at a definite temperature. On emerging from the chamber the gas is immediately passed into an absorbent which can either retain the unchanged reactant or the product or both, so that by a suitable analytical process the extent of decomposition can be determined. The procedure for calculating the velocity constant can best be explained by means of an example, and the decomposition of nitrogen pentoxide will be considered (J. K. Hunt and F. Daniels, 1925). A stream of dry nitrogen was sent through a flow-meter, and then over solid nitrogen pentoxide, the vapor of which was carried into the decomposition tube; the exit gases were then passed through standard alkali in an absorption tube. From the amount of alkali neutralized the total nitrogen pentoxide entering the reaction chamber was calculated, and from the amount of nitrite obtained the quantity of nitrogen dioxide and tetroxide in the exit gases was estimated; the latter amount gives a measure of the extent of decomposition of the pentoxide. In one experiment 0.00319 mole of pentoxide entered the decomposition chamber kept at 45° c. and 0.000126 mole was decomposed. The flow-meter indicated that during the experiment, which lasted 32 min., 1140 ml. of nitrogen, measured at 21° c. and 757 mm. pressure, had passed; in the decomposition chamber this would expand to 1194 ml., and as the nitrogen pentoxide vaporized had a volume of 83.5 ml. calculated for 0.00319 mole at 45° c. and 757 mm., the total volume of gas was 1277.5 ml. Since this quantity passed in 32 min., the rate per minute was 39.9 ml., and as the volume of the reaction chamber was 60.55 ml., each mole remained in it for $60.55/39.9 = 1.517$ min. It follows, therefore, that a is 0.00319, x is 0.000126 and t is $1.517 \times 60 = 91.0$ sec., and hence

$$k = \frac{2.303}{t} \log \frac{a}{a-x} = \frac{2.303}{91.0} \log \frac{0.00319}{0.003064} = 4.5 \times 10^{-4} \text{ sec.}^{-1}.$$

The result agrees with those in Table 172, obtained by the static method, within the limits of the relatively large experimental errors. If the reacting substance is itself a gas, or if the reaction is between two gases, so that there is no carrier gas, the quantity passing into the decomposition chamber can be determined

directly from the rate of flow without having recourse to analysis. The changes necessary in the method of calculation should be obvious.

Decomposition of Definite Fraction of Reactant.—The observations of C. N. Hinshelwood and W. K. Hutchison (1926) on the decomposition of gaseous acetone provide an illustration of another method for showing that a reaction is of the first order. It has been seen that for such reactions the time taken for a definite fraction of the reacting material to decompose is independent of the initial concentration, and this provides one of the most useful tests for a first order reaction. In the first column of Table 173 are given the initial pressures of acetone, and under the headings

TABLE 173. THERMAL DECOMPOSITION OF ACETONE AT 601° C.

Initial Pressure	$t_{0.25}$	$t_{0.5}$	$t_{0.75}$
362 mm.	26 sec.	77 sec.	199 sec.
296	27	80	200
230	24	85	240
192	20	78	—
98	30	86	202

$t_{0.25}$, $t_{0.5}$ and $t_{0.75}$ are the times, in seconds, taken for the gas pressure to increase by 25, 50 and 75 per cent respectively; they are approximately the times required to decompose these proportions of the acetone initially present. In each column the times are seen to be almost independent of the initial pressure of acetone, so that the reaction is apparently of the first order. The process is, however, not a simple one; if it were, the values of $t_{0.5}$ and $t_{0.75} - t_{0.5}$ should be the same, for they represent the times required for equal fractions of decomposition. The decomposition of acetone is, in fact, a complex reaction which probably involves a chain mechanism (cf. p. 1111).

First Order Reactions in Solution.—Certain first order reactions occurring with measurable velocity in the gas phase, at temperatures which are not too high, are of the first order also in solution. An example is the decomposition of nitrogen pentoxide which occurs with almost the same velocity at 25° c. in the gaseous state as when dissolved in nitromethane, bromine, carbon tetrachloride, chloroform and ethylene chloride (H. Eyring and F. Daniels, 1930). The isomeric change of gaseous *d*-pinene to dipentene is kinetically of the first order; the specific rate of the process is the same in the pure liquid state, and in solution in acetophenone, α -methylnaphthalene or petrolatum at 218° c. In these reactions it is evident that the solvent cannot be playing an important part. Other first order processes are known, however, for which the velocity constant depends on the nature of the solvent. The thermal decomposition of 2:4:6-trinitrobenzoic acid into carbon dioxide and trinitrobenzene, and the decomposition of triethylsulfonium bromide into diethyl sulfide and ethyl bromide are illustrations of this type of behavior. The specific rates for the former reaction in four different solvents (C. N.

TABLE 174. DECOMPOSITION OF TRINITROBENZOIC ACID IN SOLUTION

Solvent	k	Solvent	k
Water	3.33×10^{-6}	Nitrobenzene	4.07×10^{-9}
Acetophenone	5.79×10^{-7}	Toluene	1.62×10^{-7}

Hinshelwood and E. A. Moelwyn Hughes, 1931) at 60° c. are given in Table 174. Other first order reactions in which the solvent appears to be directly involved are the decomposition of ozone in carbon tetrachloride solution, and of dibromosuccinic acid into bromomaleic and hydrobromic acids in aqueous solution. The kinetics of the latter process were first studied by J. H. van't Hoff (1884), and the results quoted in Table 175

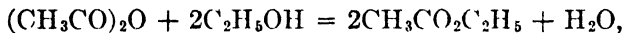
TABLE 175. DECOMPOSITION OF DIBROMOSUCCINIC ACID IN AQUEOUS SOLUTION

(a = 5.11 ml.)

t	$a - x$	k (sec. ⁻¹)
20 min.	2.74 ml.	5.17×10^{-4}
30	2.02	5.13
40	1.48	5.15
50	1.08	5.18
60	0.80	5.15

are based on his work; the course of the reaction is followed by determining volumetrically the amount of hydrobromic acid formed after definite intervals of time. If the values of $\log (a - x)$ are plotted against t , the data fall very closely on a straight line, in accordance with the requirements of a first order reaction.²

Pseudo-unimolecular Reactions.—All the reactions considered above are probably unimolecular as well as of the first order; that is to say, each act of decomposition involves one molecule only. There are, however, reactions of various kinds which satisfy the first order equations, but actually more than one molecule is involved in the chemical reaction; such processes are sometimes called **pseudo-unimolecular reactions**. The first to be considered are those in which the solvent molecules are involved chemically, as distinct from the purely physical rôle to be attributed to the solvent in the cases already discussed. A simple illustration is the interaction between acetic anhydride and ethyl alcohol, the former being in dilute solution in the latter; the reaction is



and consequently, according to the law of mass action, the velocity should be dependent on the concentrations of the two reactants; thus,

$$\frac{dx}{dt} = k c_{(\text{CH}_3\text{CO})_2\text{O}} \times c_{\text{C}_2\text{H}_5\text{OH}}^2.$$

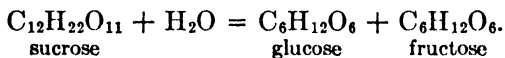
Since the alcohol is present in very large excess the amount used up in the course of the reaction is negligible in comparison with the total, so that the concentration may be regarded as remaining constant through-

out; the velocity equation then takes the form

$$\frac{dx}{dt} = k'c_{(\text{CH}_3\text{CO})_2\text{O}},$$

so that the rate is proportional to the concentration of the acetic anhydride. The process should thus be kinetically of the first order, and this has been verified by experimental observation of the reaction. The rate is followed by removing quantities of the reacting system from time to time, and determining the amount of unchanged anhydride by titration with alkali after decomposition by water. The hydrolysis of carbonyl sulfide in aqueous solution, and of substituted benzyl chlorides in alcohol-water mixtures, as well as the reaction between ethyl iodide and dimethylaniline in excess of the latter, are kinetically first order reactions although they are all actually bimolecular processes; they may, therefore, be described as pseudo-unimolecular.

Inversion of Sucrose.—If, in addition to a solvent molecule, a catalyst present in solution is also concerned in the reaction, the process may still be of the first order; an example, which has been thoroughly investigated, is the hydrolysis, or inversion, of sucrose in the presence of hydrogen ions as catalyst, viz.,



It is probable that the rate of the process depends on the concentrations of sugar, water and hydrogen ions; the concentration of water, as explained above, remains virtually constant and so also does that of the hydrogen ions, since they are not used up in the process, and so it follows that the reaction rate will be proportional to the sucrose concentration only. The inversion will consequently be kinetically of the first order.

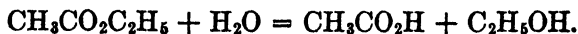
This is an example of a reaction which can be followed by measurement with a polarimeter, for the sucrose rotates the plane of polarization of light to the right, whereas the resulting mixture of glucose and fructose is *levo*-rotatory. If α_0 and α_∞ are the angles of rotation, *including the correct signs*, at the beginning and when the reaction is completed, respectively, and α_t is the value after time t , then $\alpha_0 - \alpha_\infty$ is proportional to the initial concentration a , and $\alpha_t - \alpha_\infty$ is proportional to $a - x$, the amount of unchanged sugar; hence, for a first order process,

$$k = \frac{2.303}{t} \log \frac{\alpha_0 - \alpha_\infty}{\alpha_t - \alpha_\infty}. \quad (9)$$

The applicability of this type of equation to the inversion of sucrose was first shown by L. Wilhelmly (1850) and has been amply confirmed by others; the reaction is thus pseudo-unimolecular. It should be noted that the actual value of k depends, of course, on the concentration of hydrogen ions and other catalysts present in the solution.

Hydrolysis of an Ester.—Another process of an exactly analogous type is the hydrolysis of an ester, e.g., ethyl acetate, in aqueous solution

in the presence of a mineral acid to act as catalyst; viz.,



The kinetics of the reaction are dependent only on the ethyl acetate concentration, apart from that of the catalyst, and the hydrolysis should be of the first order; this has been verified experimentally.

The disappearance of the ester can be followed by withdrawing portions of the reacting mixture from time to time and titrating with alkali. The volume of titrant required is equivalent to the sum of the acid used as catalyst, which remains constant throughout, and of the acetic acid produced in the hydrolytic reaction. The titration at the commencement of the process T_0 is equivalent to the amount of acid catalyst, and T_∞ , when the reaction is finished, to the sum of this quantity and the acetic acid resulting from complete hydrolysis of the ester; hence $T_\infty - T_0$ is a measure of a , the original ester concentration. Similarly $T_\infty - T_t$, where T_t is the titration after time t , gives the quantity of ester $a - x$ remaining at that time. It follows, therefore, for a first order process, that

$$k = \frac{2.303}{t} \log \frac{T_\infty - T_0}{T_\infty - T_t}, \quad (10)$$

so that the specific rate can be evaluated.

Other Reactions.—Another type of pseudo-unimolecular reaction is that between a single molecule of reactant and a catalyst, without the solvent being involved; since the concentration of the catalyst remains constant such reactions must be kinetically of the first order. A good illustration is the conversion of *N*-chloroacetanilide to the isomeric *p*-chloroacetanilide in the presence of hydrochloric acid as catalyst, i.e., $\text{C}_6\text{H}_5\text{NClCOCH}_3 \rightarrow \text{ClC}_6\text{H}_4\text{NHCOCH}_3$ (cf. p. 177); the progress of the reaction is readily followed by utilizing the fact that the *N*-chloro-compound liberates iodine from potassium iodide. The volume of thiosulfate solution required to titrate the iodine is a direct measure of the concentration of unchanged reactant. In Table 176 are quoted some

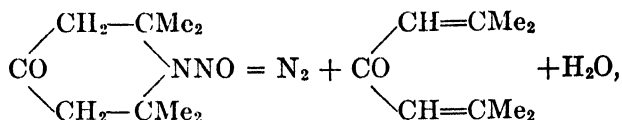
TABLE 176. ISOMERIC CHANGE OF *N*-CHLOROACETANILIDE

t	$(a = 24.5 \text{ ml. Na}_2\text{S}_2\text{O}_3 \text{ solution})$ $a - x$	k
15 min.	18.1 ml.	0.0202 min. ⁻¹
30	13.3	0.0204
45	9.7	0.0206
60	7.1	0.0206
75	5.2	0.0207

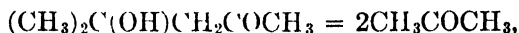
results obtained in this manner by J. J. Blanksma (1902); the values of k given by (4) serve to show that the isomeric change is undoubtedly of the first order.

Two other reactions of the same type, involving catalyst and one molecule of reacting substance, are of interest. One is the decomposition

of nitrosotriacetoneamine into nitrogen, phorone and water, viz.,

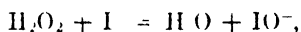


in the presence of hydroxyl ions as catalyst; the rate can be followed by measuring the volume of nitrogen evolved after definite intervals of time. The second reaction is the conversion of diacetone alcohol to acetone, that is



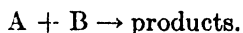
and this is also catalyzed by hydroxyl ions. The specific rate has been determined by observing the change in volume by means of a dilatometer, for the reaction is accompanied by an increase of volume. The isomeric change of α - to β -glucose, generally referred to as the mutarotation of glucose, is a reaction of the same type; as it is a reversible one, however, it should be treated by the methods used in connection with such processes (p. 1069).

Decomposition of Hydrogen Peroxide.—The catalytic decomposition of hydrogen peroxide in the presence of finely divided platinum and of the enzyme blood catalase are often quoted as first order reactions. It is certainly true that they are found to be so kinetically, but the reactions are heterogeneous. Even in a pure aqueous solution it is not certain how far the decomposition of the peroxide is influenced by the walls of the containing vessel or by dust particles. The reaction in the presence of iodide ions as catalyst in neutral solution is, however, undoubtedly homogeneous, and although the process is not as simple as might at first appear, it seems definite that the essential step determining the rate of the reaction involves one molecule only of hydrogen peroxide; thus,



the IO^- ions being subsequently reconverted into iodide ions. In agreement with this view, the decomposition has been found to be kinetically of the first order; the actual specific rate, for a given iodide concentration, depends on the concentration of electrolytes in the solution, as is frequently the case with reactions involving ions (see p. 1115). The best method for following the course of the decomposition is to measure the volume of oxygen gas (x) evolved from time to time, and also when the reaction is complete (a); the data when inserted in the first order equation give very satisfactory constants.

Second Order Reactions.—When the velocity of a reaction depends on two concentration terms the process is said to be one of the second order. If two reactant molecules A and B are concerned, then the simplest second order reaction, which is also bimolecular, may be represented by



If a and b are the initial concentrations of A and B respectively, and x is the decrease of each after time t , then since the rate of reaction dx/dt is

proportional to the concentration of both reactants, it follows that

$$\frac{dx}{dt} = k(a-x)(b-x). \quad (11)$$

On integration, taking into consideration that $x = 0$ when $t = 0$, and that $x = x$ when $t = t$, it is seen that

$$k = \frac{1}{t(a-b)} \ln \frac{b(a-x)}{a(b-x)} = \frac{2.303}{t(a-b)} \log \frac{b(a-x)}{a(b-x)}, \quad (12)$$

where k is the second order reaction constant. An equation of this type was apparently first deduced by M. Berthelot (1862) in connection with his study of the reaction between ethyl alcohol and acetic acid.

If a and b are not very different then the fraction $b(a-x)/a(b-x)$ in (12) is very close to unity, and unless seven figure logarithms are employed the results will not be accurate. Under these conditions it is perhaps better to adopt a somewhat different treatment: equation (12) may be written

$$k = \frac{1}{t(a-b)} \ln \left[\frac{a-(a-b)}{a} \cdot \frac{a-x}{a-x-(a-b)} \right] \quad (13)$$

$$= \frac{1}{t(a-b)} \left[\ln \left(1 - \frac{a-b}{a} \right) - \ln \left(1 - \frac{a-b}{a-x} \right) \right] \quad (14)$$

If the logarithmic terms are expanded as series, it follows that

$$k = \frac{1}{t(a-b)} [(A + \frac{1}{2}A^2 + \frac{1}{3}A^3 + \dots) - (B + \frac{1}{2}B^2 + \frac{1}{3}B^3 + \dots)], \quad (15)$$

where A is equal to $(a-b)/(a-x)$ and B is $(a-b)/a$. This form of the velocity equation may be used if the difference between a and b is not large.

In the limiting case when $a-b$ is extremely small, so that terms involving A^2 , A^3 , \dots , B^2 , B^3 , \dots , etc. can be neglected, (15) becomes

$$k = \frac{1}{t(a-b)} (A - B) = \frac{1}{t} \cdot \left(\frac{1}{a-x} - \frac{1}{a} \right) \quad (16)$$

or

$$k = \frac{1}{t} \cdot \frac{x}{a(a-x)}, \quad (17)$$

and this is the form applicable when a and b are equal, that is when the two reacting substances are at the same concentration. The same result is obtained by putting $a = b$ in (11), which then becomes

$$\frac{dx}{dt} = k(a-x)^2, \quad (18)$$

and on integration (17) follows directly. The simplified equations are, of course, applicable when the reaction involves two molecules of the same substance; thus,



for in these circumstances a and b are obviously identical.

Examination of (12) and (17) shows that the value of the second order specific rate differs from that for a first order reaction in being dependent on the units employed to express concentrations; it can be seen that there is always one more concentration term in the denominator than in the numerator, so that the velocity constant is expressed in $\text{conc.}^{-1} \text{ time}^{-1}$ units. A common form is $(\text{mole/liter})^{-1} \text{ sec.}^{-1}$ or $\text{liter mole}^{-1} \text{ sec.}^{-1}$, the concentration being in moles per liter and the time in seconds.

The time taken for the initial concentrations to be reduced by a definite fraction may be related to the concentration by (17). For example, the time $t_{0.5}$ for the initial amount to be reduced by half is given by substituting $a/2$ for x , when it is seen that

$$t_{0.5} = \frac{1}{ka}. \quad (19)$$

The time is then inversely proportional to the initial concentration; the same relationship applies for any other definite fraction. It is thus possible to distinguish in a simple manner between first and second order reactions, especially when only one reacting substance is involved.

In addition to this method and the direct one of determining whether (12) or (17) is applicable, according to circumstances, graphical devices may be used to show that a reaction is of the second order. If the process involves two reactants at different concentrations, then (12) may be re-arranged to take the form

$$t = \frac{2.303}{k(a-b)} \log \frac{b}{a} + \frac{2.303}{k(a-b)} \log \frac{a-x}{b-x}, \quad (20)$$

and since a and b are constants for a given reaction mixture, it is clear that the plot of t against $\log (a-x)/(b-x)$ should be a straight line, the slope of which is equal to $2.303/k(a-b)$. If the initial concentrations of the reacting substances are the same, or two molecules of only one reactant take part in the change, then (16), which applies under these conditions, may be written

$$t = \frac{1}{k(a-x)} - \frac{1}{ka}, \quad (21)$$

so that the plot of t against $1/(a-x)$ should be linear, the slope being $1/k$.

It has been stated above that if one of the reactants in a bimolecular process is present in excess, the reaction becomes kinetically of the first order; this can be readily shown by making the assumption in (12) that b is large in comparison with a and x , so that $a-b \approx -b$, and $b-x \approx b$.

The equation then becomes

$$k = \frac{1}{tb} \ln \frac{a}{a-x},$$

and since b is constant this is equivalent to (4), for a first order reaction.

Gas Reactions of the Second Order.—The thermal decomposition of acetaldehyde vapor is probably a complicated process involving various mechanisms; nevertheless, over a certain range of temperature and pressure the reaction appears to be of the second order, as is shown by the data in Table 177 (C. N. Hinshelwood, 1926). The amount of aldehyde

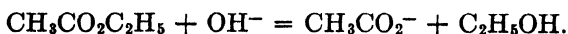
TABLE 177. THERMAL DECOMPOSITION OF ACETALDEHYDE AT 518° C.
(Initial pressure $a = 363$ mm.)

t	$\Delta p = x$	k
42 sec.	34 mm.	6.79×10^{-8}
105	74	6.71
242	134	6.66
480	194	6.59
840	244	6.72
1440	284	6.86

decomposed (x) is determined by the increase of pressure (Δp), since one molecule on decomposition should give two molecules of products, viz., methane and carbon monoxide. In the determination of k the concentrations were expressed in terms of mm. of mercury pressure, and (17) was used. The values in the last column show that the decomposition of acetaldehyde under the given conditions is apparently a reaction of the second order; it is probably not a simple bimolecular process, for it involves a chain mechanism (p. 1084), and an order of 1.5 has been observed. Other homogeneous gas reactions which have been said to be, at least partly, of the second order are the thermal decomposition of nitrogen dioxide, chlorine monoxide and ozone; these reactions are, however, more complex than appears at first sight. The combination of ethylene and hydrogen and the polymerization of ethylene are also of the same order. It will be seen later that as a general rule unimolecular first order gas reactions become kinetically of the second order at low pressures; this result is of great theoretical significance. It should be emphasized that many gas reactions which are of the second order appear to be complicated by the existence of a number of concurrent reactions, and consequently great caution has to be used in interpreting the results. A few cases of second order reactions apparently free from serious complication are known; of these the most important are the combination of hydrogen and iodine and the reverse reaction, the decomposition of hydrogen iodide. The two processes, however, occur simultaneously and the results must be treated by special methods adopted for reversible reactions (p. 1072).

Reactions in Solution—One of the earliest reactions of the second order occurring in solution to be studied was that between ethyl acetate

and alkali hydroxide, viz.,



It differs from the hydrolysis in acid solution considered above, because neither reactant need be present in excess, and so it can be of the second order. The course of the reaction can be followed by removing definite quantities of the reacting system from time to time and running into excess of standard acid; the excess is then titrated with alkali, so that the concentration of unchanged hydroxide in the reaction mixture can be determined.

If T_0 is the volume of standard acid required to neutralize the hydroxide at the beginning of the experiment, T_t is the value after t sec. and T_∞ that equivalent to the excess alkali remaining when all the ester has been saponified, then (12) can be written

$$k = \frac{2.303}{tT_\infty} \log \frac{T_t(T_0 - T_\infty)}{T_0(T_t - T_\infty)}. \quad (22)$$

In order to express the results in terms of liter mole⁻¹ sec.⁻¹ units, it is necessary to divide the value of k found in this way by n/v , where v is the volume of reaction mixture in ml. withdrawn for each titration and n is the normality of the acid used; this is the factor for converting titrations into concentrations in g.-equiv. per liter, which in this case are identical with moles per liter.

Some of the results obtained by L. T. Reicher (1887) in this manner are given in Table 178. Reactions between other esters and alkali

TABLE 178. SAPONIFICATION OF ETHYL ACETATE BY SODIUM HYDROXIDE AT 15.8° C.
(Acid = 0.04155 N; v = 100.2 ml.)

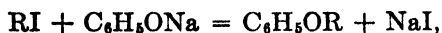
t sec.	T_t ml.	k (liter mole ⁻¹ sec. ⁻¹)
0	62.09 (= T_0)	
224	54.33	5.77×10^{-2}
377	50.60	5.79
629	46.28	5.71
816	43.87	5.72
24 hrs.	33.06 (= T_∞)	—

hydroxides have been examined in a similar manner. The rate of a saponification reaction may also be studied by measuring the conductance of the reaction mixture after definite time intervals; the decrease of conductance, resulting from the replacement of hydroxyl by acetate ions, is approximately proportional to the decrease of concentration (J. Walker, 1906).

Other second order reactions taking place in solution are the interaction of alkyl halides with tertiary amines or pyridine, e.g., ethyl iodide and dimethylaniline

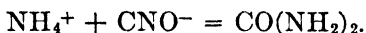


which has been studied in various solvents, and of alkyl iodides with sodium phenoxide, in alcohol solution, e.g.,

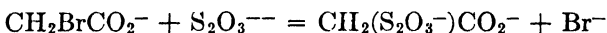


where R varies from CH_3 to C_8H_{17} . The progress of both these types of process can be followed by titration, with silver nitrate, of the iodide ions produced in the reaction.

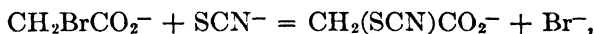
It is often assumed that ionic reactions are so rapid that their velocities cannot be determined without the use of special devices, but this is not always true; there are a number of processes involving ions only but in which covalent links are broken and others formed, and these take place with a velocity which is measurable at ordinary temperatures. The rate of isomeric conversion of ammonium cyanate to urea in aqueous solution has been measured both by titration of the cyanate and by determining its conductivity (J. Walker, 1895). The process is of the second order and involves interaction between two ions; thus



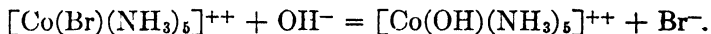
The reaction is really reversible, but it goes almost to completion from left to right in dilute solution, and the reverse change can be ignored. Other relatively slow reactions in which two ions are concerned are the reactions in aqueous solution between bromoacetates and either thiosulfate or thiocyanate ions, viz.,



and



and that between bromopentammine cobaltic and hydroxyl ions, viz.,



These ionic reactions are of special interest, as will be seen later (p. 1115); their velocities are dependent on the total concentration of electrolyte in the solution.³

Pseudo-bimolecular Reactions.—Just as there are bimolecular reactions which are kinetically of the first order because one of the reactants is present in excess, so there are processes involving three molecules which are of the second order. An example of this type of reaction is found in the esterification which occurs when a carboxylic acid (RCO_2H) is dissolved in an alcohol ($\text{R}'\text{OH}$), the molecules of acid acting as catalyst; the rate of reaction is then given by

$$\frac{dx}{dt} = k c_{\text{RCO}_2\text{H}} \times c_{\text{R}'\text{OH}} \times c_{\text{RCO}_2\text{H}}, \quad (23)$$

and since the alcohol is present in excess, this reduces to

$$\frac{dx}{dt} = k' c_{\text{RCO}_2\text{H}}^2 \quad (24)$$

as required for a second order reaction. If excess of hydrogen chloride is present the rate of reaction is increased because the hydrogen ions act catalytically, and although three species are involved in the process, viz.,

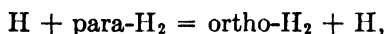
carboxylic acid, alcohol and hydrogen ions, the process is of the first order, because the concentrations of two of the reactants, viz., alcohol and hydrogen ions, are virtually constant. The reaction in which the carboxylic acid is the catalyst now takes place to such a negligible extent that it has no appreciable influence on the kinetics.

Bimolecular Reactions of Fractional Order.—The high temperature conversion of parahydrogen to the ortho-form (p. 94), and the analogous change of ortho- to para-deuterium, present a number of unusual features. The time required for half the initial amount to be converted * is found to be inversely proportional to the square root of the initial pressure, i.e., $t_{0.5} \propto 1/p^{1/2}$; this would imply a reaction of the order 1.5 [cf. equation (42)]. On the other hand, at a particular pressure, the rate of conversion satisfies the equation for a first order reaction, but the specific rate k' varies with the hydrogen pressure, which remains unchanged, of course, during the course of the reaction. However, the values of $k'/p_{H_2}^{1/2}$, where p_{H_2} is the pressure of the hydrogen, are found to be constant, as may be seen from the results in Table 179, which were obtained for the conversion of para- to ortho-hydrogen at 650° c. (A. Farkas, 1930).

TABLE 179. CONVERSION OF PARA- TO ORTHO-HYDROGEN AT 650° C.

p_{H_2}	k' (sec. ⁻¹)	$k'/p_{H_2}^{1/2}$
50 mm.	1.06×10^{-3}	1.50×10^{-4}
100	1.53	1.53
200	2.17	1.54
400	3.10	1.55

These curious results can be readily explained by the suggestion that the conversion mechanism actually involves a "bimolecular" process, between a molecule of parahydrogen and a hydrogen atom, produced by the dissociation of a molecule, viz.,



so that

$$\frac{dx}{dt} = kp_H p_{\text{para-}H_2} \quad (25)$$

expressing concentrations in terms of partial pressures. Since hydrogen atoms and molecules, both ortho- and para-, are in equilibrium, i.e., $H_2 \rightleftharpoons 2H$, it follows from the law of equilibrium, assuming ideal behavior, that $p_H^2 = Kp_{H_2}$. Introducing this result into (25), it is seen that

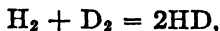
$$\frac{dx}{dt} = k_1 p_{H_2}^{1/2} p_{\text{para-}H_2}, \quad (26)$$

which corresponds kinetically to a total order of 1.5. At a given value of the total hydrogen pressure p_{H_2} , the rate will be proportional to $p_{\text{para-}H_2}$, and hence the reaction should behave as one of the first order, with the specific rate k' equal to $k_1 p_{H_2}^{1/2}$. As the pressure of hydrogen is varied,

* The quantities refer to the excess over the equilibrium ortho-para ratio in each case; by using this device the reverse reaction can be neglected (cf. p. 1070).

however, $k'/p_H^{\frac{1}{2}}$, should give the constant value k_1 , as is actually the case (Table 179).

Another "bimolecular" reaction of apparent order 1.5 is that between hydrogen and deuterium,



for working with mixtures of the two gases at equal partial pressures the time taken to complete half the reaction is inversely proportional to the square root of the initial pressure. The process probably takes place in stages; thus,



which are rapid, followed by



the hydrogen and deuterium atoms either reacting with each other or combining in pairs. The over-all rate of the reaction is evidently determined by the stages involving an atom and a molecule, for these would be kinetically of the order 1.5; the other changes are probably so rapid that they have no effect on the kinetics (cf. p. 1065).

Autocatalytic Reactions.—In some reactions one of the products is able to catalyze the further decomposition of the reactants; such a process is said to be **autocatalytic**, and a special type of second order reaction arises in this manner. For example, in the hydrolysis of an ester, e.g., methyl acetate, the acetic acid produced is able to catalyze the reaction. Let a be the concentration of ester and b of acetic acid present initially, and let x be the amount of ester hydrolyzed after time t ; then the rate of hydrolysis at any instant is given by

$$\frac{dx}{dt} = k(a - x)(b + x), \quad (27)$$

assuming the rate to be proportional to the concentration of the acetic acid catalyst. On integration in the usual way, this gives

$$k = \frac{1}{t(a + b)} \ln \frac{a(b + x)}{b(a - x)}, \quad (28)$$

which has been confirmed experimentally by W. Ostwald (1883). If the catalyst present initially is different from the one produced in the reaction, the rate of the whole process will be

$$\frac{dx}{dt} = k_1(a - x)b + k_2(a - x)x \quad (29)$$

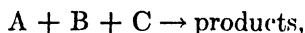
$$= (k_1b + k_2x)(a - x), \quad (30)$$

where b is the amount of catalyst originally present and x is that produced by the reaction. Integration gives

$$k_1b + k_2a = \frac{1}{t} \ln \frac{a(k_1b + k_2x)}{k_1b(a - x)}, \quad (31)$$

which has also been confirmed by Ostwald in the hydrolysis of methyl acetate with α -oxyisobutyric acid as the catalyst present initially.

Third Order Reactions.—If the rate of a reaction is determined by the variation of three concentration terms it is a process of the third order, and three different cases may arise, (i) the three terms may be equal, (ii) two may be equal and one different, or (iii) all three may be different. In the latter circumstance, e.g., for a reaction



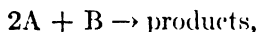
with the reactants at initial concentrations a , b and c , respectively, it follows that

$$\frac{dx}{dt} = k(a - x)(b - x)(c - x), \quad (32)$$

where x is the concentration change after time t from the commencement. Integration of this equation gives

$$k = \frac{1}{t} \cdot \frac{(b - c) \ln \frac{a - x}{a} + (c - a) \ln \frac{b - x}{b} + (a - b) \ln \frac{c - x}{c}}{(a - b)(b - c)(c - a)}. \quad (33)$$

If two of the reactants are at the same concentration, for example as would be the case for a process of the type



then

$$\frac{dx}{dt} = k(a - 2x)^2(b - x), \quad (34)$$

since the amount of A decomposed at any instant, viz., $2x$, is twice that of B, i.e., x . The integrated form is then

$$k = \frac{1}{t} \cdot \frac{1}{(a - 2b)^2} \left\{ \frac{2x(2b - a)}{a(a - 2x)} + \ln \frac{b(a - 2x)}{a(b - x)} \right\}. \quad (35)$$

In the simplest case, when reactants are all at the same concentration, or if the reaction is of the type



then

$$\frac{dx}{dt} = k(a - x)^3 \quad (36)$$

and

$$k = \frac{1}{2t} \left\{ \frac{1}{(a - x)^2} - \frac{1}{a^2} \right\}. \quad (37)$$

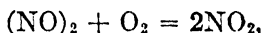
Equation (33) may be converted into (37) by the method described in connection with second order reactions, the logarithmic terms being replaced by their equivalent series; the procedure need not be repeated here. Further, if either one or two of the reacting species are assumed to

be in excess, (33) will reduce to the appropriate forms for second or first order reactions, respectively.

It will be evident that the value of the specific rate depends on the concentration units, for it is given by $\text{conc.}^{-2} \text{ time}^{-1}$, since there are always two more concentration terms in the denominator than in the numerator. The time required to complete a definite fraction of a third order reaction is inversely proportional to the square of the initial concentration, as may be seen by writing x equal to $a/2$ in (37), which then gives

$$t_{0.5} = \frac{3}{2ka^2}. \quad (38)$$

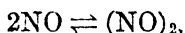
Third Order Gas Reactions.—Only five homogeneous gas reactions are definitely known to be of the third order, and these all involve nitric oxide, the other reactant being a molecule of chlorine, bromine, oxygen, hydrogen or deuterium. The fact that nitric oxide contains an odd number of electrons has led to the suggestion that something equivalent to an associated molecule $(\text{NO})_2$ is first formed and that this combines with the other reactant, in what is virtually a bimolecular process, viz.,



so that

$$\frac{dx}{dt} = kc_{(\text{NO})_2} \times c_{\text{O}_2}. \quad (39)$$

If there is an equilibrium between simple and associated nitric oxide, then



and by the law of equilibrium,

$$c_{(\text{NO})_2} = Kc_{\text{NO}}^2. \quad (40)$$

If this value for $c_{(\text{NO})_2}$ is inserted in (39) it is seen that

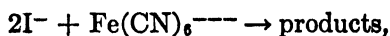
$$\frac{dx}{dt} = kKc_{\text{NO}}^2 \times c_{\text{O}_2} = k'c_{\text{NO}}^2 \times c_{\text{O}_2}, \quad (41)$$

as required for a third order process. According to this view the reactions involving nitric oxide are really bimolecular, although they are kinetically of the third order.

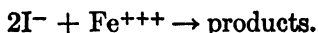
Third Order Reactions in Solution.—The reaction between stannous and ferric salts in aqueous solution is often described as being of the third order, but it is probable that this result is incorrect. Reactions involving ions are greatly influenced by the total concentration of electrolytes (p. 1115), and this factor is partly responsible for the erroneous conclusions drawn by earlier workers. More recent investigations indicate that the rate-determining stage is bimolecular and second order.

It is possible that the interaction of ferricyanide and iodide ions may

be of the third order, viz.,



as suggested by A. von Kiss (1933). This reaction was at one time considered to be of the fifth order, but here also it appears that the results were affected by the total ionic concentration of the solution. Another reaction occurring in solution which may be of the third order is that between ferric and iodide ions, i.e.,

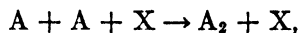


The condensation of benzaldehyde to benzoin in the presence of cyanide ions as catalyst, viz.,

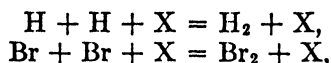


is probably termolecular, but of the second order since the amount of CN^- remains constant.³

Three-Body Collisions.—Although it was at one time thought that two atoms of an element such as hydrogen, chlorine, bromine, etc. would unite very rapidly to form the corresponding molecule, it is now known that this is not the case. The heat of reaction is so large that unless the resultant molecule can be relieved of some of this excess energy in a very short time, it will dissociate into atoms again. Most of this energy is removed by the walls of the vessel when two atoms meet there, but there is also a possibility that some of the energy will be removed when a three-body collision occurs between the two atoms and a molecule capable of taking off some of the energy of the reaction. The homogeneous recombination of atoms is thus a termolecular process, the term "molecular" including atomic reactants, viz.,



where A represents the combining atoms and X is the molecule capable of removing energy. Since the concentration of X remains unchanged, the reaction is kinetically of the second order. Three-body collisions of the type under consideration, sometimes involving atoms, e.g.,

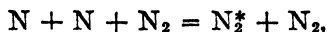


or an unstable radical, e.g.,



are of considerable importance in many chemical reactions (cf. pp. 1172, 1175). An interesting illustration is provided by the decay of active nitrogen; a study of the spectrum of the glow emitted shows that it is produced by excited nitrogen *molecules* returning to the ground state.

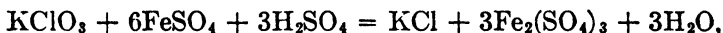
After passage through the electric discharge, the active nitrogen contains a proportion of atoms, and these apparently recombine in pairs in a ternary collision with a nitrogen molecule, thus



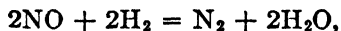
the energy of the recombination process being sufficient to raise the nitrogen molecule to an excited level. This explanation is in harmony with the following facts: the active nitrogen decays, at constant total pressure, in a bimolecular manner with respect to nitrogen atoms; the rate of decay is proportional to the total pressure of the nitrogen, that of the atoms being very small in comparison; and it takes place more rapidly at low than at high temperatures, as is sometimes found for termolecular reactions (see p. 1106).⁴

Reactions of Higher Orders.—In the literature there will be found references to reactions of orders higher than three, even as high as eight; it is very unlikely, however, that there are any processes higher than of the third order. A reaction of the fourth order would mean that four molecules of the correct kind would have to meet simultaneously before interaction could occur; the probability of this taking place would be very small. Effective collisions between three molecules are not very common, as is evident from the limited number of third order reactions, and it is probable that fourth order reactions do not occur. A quadrimolecular reaction involving one or two solvent molecules, or other substance, present in large excess, is not impossible, but these would be kinetically of a lower order than the fourth. Higher orders have been found experimentally for ionic reactions, and it seems that in such cases the effect of the electrolyte has been a disturbing factor. There is also another reason for the apparent high order which will be explained later (p. 1069).

Complex Reactions.—Since the probability that more than three molecules, or perhaps four if one is present in large excess, can meet and react simultaneously is very small, it is evident that complex reactions involving more than three molecules in the stoichiometric equation, e.g.,



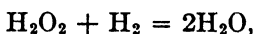
must take place in stages. Consideration will show that it is the slowest stage which determines the over-all rate of the whole process, especially if all other stages occur with great rapidity. The above reaction, which is apparently of the tenth order, is actually of the second order, so that it is probable there is at least one slow stage involving collision between two molecules. The complete reaction between nitric oxide and hydrogen is



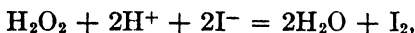
although kinetically the process is of the third order; this result suggests that the reaction takes place in two stages, viz., a slow stage



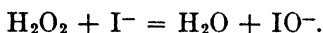
which is the rate-determining process, followed by



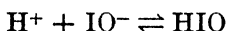
taking place so rapidly that it has no influence on the kinetics of the reaction. The oxidation of hydriodic acid in aqueous solution by hydrogen peroxide is written



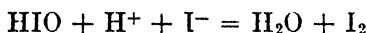
but as it is found to be of the second order, the rate-determining slow stage is probably



Since hypiodous acid is very weak, the equilibrium



is established instantaneously, and then a rapid reaction such as



may follow. The whole oxidation process thus involves at least three steps, two of which being very fast do not affect the kinetics. The rate of formation of iodine, or of the disappearance of hydrogen peroxide, is determined almost exclusively by the first, i.e., slow, stage, and the order found experimentally will be the order of this stage. It will be seen from the examples quoted that a knowledge of the order of a reaction can often be used to throw light on the mechanism of the process; this is one of the important applications of reaction kinetics in chemistry.

Determination of Order of Reaction.—It will be opportune now to review the various methods available for determining the order of a reaction.

(i) *Method of Integration.*—The most obvious procedure is to determine by analysis the amount of reactant decomposed after various intervals of time, and to substitute the data in the various equations for reactions of the first, second and third order. The equation giving the most constant value for the specific rate for a series of time intervals is the one most closely corresponding to the correct order of the reaction.

(ii) *Time to Complete Definite Fraction of the Reaction.*—It has been seen that the time taken to complete a definite fraction, e.g., one half, of the reaction, provided the reactants are all at the same initial concentration, is independent of the initial concentration for a process of the first order, inversely proportional to it for a second order reaction, and inversely proportional to its square for a third order process. This result may be generalized in the form

$$\tau \propto \frac{1}{a^{n-1}} \quad (42)$$

where τ is the time required to complete a certain fraction of the reaction,

a is the initial concentration of reactants and n is the order of the reaction. If in two different experiments the concentrations are a_1 and a_2 and the corresponding times are τ_1 and τ_2 , then

$$\frac{\tau_1}{\tau_2} = \left(\frac{a_2}{a_1} \right)^{n-1}; \quad (43)$$

$$\therefore n = 1 + \frac{\log \tau_1 / \tau_2}{\log a_2 / a_1}, \quad (44)$$

so that n can be evaluated directly. Examples of the application of this method, which was first used by W. Ostwald (1888), have been given in Tables 173 and 179.

(iii) *The Differential Method.*—This method, suggested by J. H. van't Hoff (1884), depends on the fact that the rate of a reaction of the n th order is proportional to the n th power of the concentration, viz.,

$$-\frac{dc}{dt} = kc^n, \quad (45)$$

where c is the concentration at any instant. For two different concentrations,

$$-\frac{dc_1}{dt} = kc_1^n \quad \text{and} \quad -\frac{dc_2}{dt} = kc_2^n, \quad (46)$$

and by taking logarithms it is easily shown that

$$n = \frac{\log (-dc_1/dt) - \log (-dc_2/dt)}{\log c_1 - \log c_2}. \quad (47)$$

It should be noted that the result will be independent of the units employed to express the concentrations.

If the reaction involves more than one substance it is not necessary that they should all have the same initial concentration; for example, if there are two reactants A and B, then

$$-\frac{dc}{dt} = kc_A^{n_A} \times c_B^{n_B}, \quad (48)$$

where n_A and n_B are the orders with respect to the two reactants, and the total order is $n_A + n_B$. If in two experiments c_A is varied but c_B is kept constant, it is clear that the application of (47) will give the value of n_A ; on the other hand if c_A is maintained constant in two experiments and c_B varied, n_B may be evaluated.

An approximate method may be used to determine dc/dt : the rate of change in c over an appreciable time interval, i.e., $\Delta c/\Delta t$, is measured and assumed to be dc/dt corresponding to the mean value of c in the interval considered. From the work of L. T. Reicher (1895) on the action of bromine on fumaric acid, the following data were obtained:

Experiment I			Experiment II		
t	c	$-\Delta c/\Delta t$	t	c	$-\Delta c/\Delta t$
0	8.88	0.0106	0	3.81	0.00227
95 min.	7.87		132 min.	3.51	

Mean concentration = 8.37

Mean concentration = 3.66

Substituting 8.37 for c_1 and 0.0106 for $-dc_1/dt$, and 3.66 and 0.00227 for c_2 and $-dc_2/dt$, respectively, in (47), n is found to be 1.87, and so the reaction must be of the second order. -

A better procedure to be adopted when possible is to plot the values of c for various times t , and to measure the slopes of the tangents at several places; in this way a more accurate estimate of dc/dt can often be made. In illustration of this method the data on the reaction between nitric oxide and hydrogen (C. N. Hinshelwood, 1926) may be employed. When the hydrogen pressure was kept constant and that of the nitric oxide varied, the following observations were made:

p_{NO}	$-dp/dt$	n
359 mm.	1.50 mm./sec.	} 2.1
300	1.03	
152	0.25	} 2.1

and when the pressure of the nitric oxide was constant, the results were

p_{H_2}	$-dp/dt$	n
289 mm.	1.60 mm./sec.	} 1.1
205	1.10	
147	0.79	} 1.0

The values of n have been calculated with the aid of (47), concentrations being expressed by the pressure of the gas. It is evident that the reaction is of the second order with respect to nitric oxide and of the first order with regard to hydrogen; the whole process is therefore of the third order.

(iv) *The Isolation Method.*—If all the reactants except one are taken in excess, it is clear that the observed order will be determined by the number of molecules of the reactant isolated in this way that take part in the reaction. The principle of isolation has been used for many years, but the method for determining the order of a process with several reactants, by isolating each in turn, is generally attributed to W. Ostwald (1902). Suppose a reaction involves n_A molecules of A, n_B of B, and n_C of C; then the total order will be $n_A + n_B + n_C$. If B and C are taken in excess the process will be of the order n_A , which can be determined by any of the three methods just described. Similarly n_B and n_C can be evaluated by studying the kinetics of the reaction with A and C, and A and B, respec-

tively, in excess. It is necessary to point out, however, that the results obtained should be accepted with caution. For example, in the reaction between iodide and iodate ions in acid solution, which is actually a complicated process, the isolation procedure suggests that at least part of the reaction is determined by the expression

$$\frac{dx}{dt} = kc_{\text{IO}_3^-} \times c_{\text{I}^-}^2 \times c_{\text{H}^+}^2 \quad (49)$$

so that the whole is of the fifth order. It is generally thought, however, that such a reaction is improbable, and the conclusion must be drawn that the interpretation of the facts is uncertain. In many instances the isolation method gives the order of the reaction correctly, especially if there are only two reactants, but in other cases the results obtained may be misleading.

SIMULTANEOUS REACTIONS

The treatment hitherto has been restricted to reactions which are relatively simple and direct; it is now necessary to consider the complications arising from the possibility of the occurrence of other processes at the same time, e.g., opposing, consecutive or side reactions. Such complexities frequently occur in practice, although the extent of the simultaneous reactions is often small in comparison with the main reaction and so can be neglected.

Opposing Reactions.—When a process is reversible, in the chemical sense, the direct and reverse reactions must occur simultaneously, and if the latter is appreciable it must be taken into account. The simplest case to consider is that in which both reactions are of the first order; thus,



If a is the concentration of A at the commencement of the experiment, and there is initially no B present, then after time t the concentrations of A and B will be $a - x$ and x , respectively, so that the net rate of reaction at any instant will be given by

$$\frac{dx}{dt} = k(a - x) - k'x, \quad (50)$$

where k is the specific rate of the forward reaction and k' that of the opposing reaction. When the system reaches equilibrium the rates of the two reactions are equal, so that

$$k(a - x_e) = k'x_e, \quad (51)$$

where x_e is the amount of B formed or of A changed at equilibrium.

Substituting the value for k' obtained from (51) in (50), it follows that

$$\frac{dx}{dt} = k(a - x) - \frac{kx}{x_e}(a - x_e) = \frac{ka}{x_e}(x_e - x), \quad (52)$$

which on integration, noting that $x = 0$ when $t = 0$ and $x = x_e$ when $t = t$, gives

$$\frac{ka}{x_e} = \frac{1}{t} \ln \frac{x_e}{x_e - x}. \quad (53)$$

Hence the specific rate of the direct reaction can be evaluated if a and x_e are known. An alternative equation is often used, however, in the study of reversible reactions. It can easily be shown by means of (51) that ak/x_e is equal to $k + k'$, so that (53) becomes

$$k + k' = \frac{1}{t} \ln \frac{x_e}{x_e - x}. \quad (54)$$

It is seen, therefore, that a reversible first order reaction may be treated as if it were not reversible, the initial concentration (a) being replaced by the amount changed at equilibrium (x_e); the specific rate so obtained, however, is the sum of those for direct and reverse reactions. An example of the applicability of (54) is taken from the work of T. M. Lowry (1899) on the mutarotation of π -bromonitrocarnphor in chloroform solution at 14°C. , which is a reaction of the type under discussion; the change is followed polarimetrically and the method of calculation is similar to that described for the inversion of cane sugar. In the present instance the appropriate form of (54) is

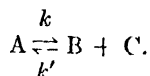
$$k + k' = \frac{1}{t} \ln \frac{\alpha_0 - \alpha_t}{\alpha_t - \alpha_e}, \quad (55)$$

where α_0 , α_t and α_e are the angles of rotation at the commencement, after time t , and at equilibrium, respectively. The results are recorded in Table 180. The constancy of $k + k'$ in the last column shows that (54) and (55) represent the kinetics of the mutarotation.

TABLE 180. MUTAROTATION OF π -BROMONITROCAMPBOR IN CHLOROFORM SOLUTION AT 14°C.

t	$(\alpha_0 = 189^\circ \quad \alpha_e = 31.3^\circ)$ α_t	$k + k' \text{ (sec.}^{-1}\text{)}$
3 hrs.	169.0°	1.26×10^{-5}
5	156.0	1.27
7	146.0	1.27
24	84.5	1.26
72	37.3	1.26

The next case to consider is a first order reaction opposed by one of the second order; thus,



Assuming that no B and C are present at the commencement of the process, but accumulate as it proceeds, the net rate of the forward reaction is given by

$$\frac{dx}{dt} = k(a - x) - k'x^2, \quad (56)$$

where $a - x$ is the concentration of A at any instant and x is that of both B and C. At equilibrium the direct and reverse processes have equal velocities so that

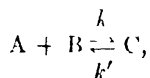
$$k(a - x_e) = k'x_e^2. \quad (57)$$

Substituting the value of k' obtained from (57) in (56), and integrating in the usual manner, after splitting into partial fractions, the result is

$$k = \frac{x_e}{t(2a - x_e)} \ln \frac{ax_e + x(a - x_e)}{a(x_e - x)}. \quad (58)$$

An equation of this type has been found to apply in the decomposition of certain alkylammonium halides into a tertiary amine and an alkyl halide in solution.

If a direct second order reaction is opposed by one of the first order, viz.,



then the simplest case arises when A and B have the same concentration but no C is present initially. It then follows that

$$\frac{dx}{dt} = k(a - x)^2 - k'x, \quad (59)$$

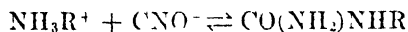
and at equilibrium

$$k(a - x_e)^2 = k'x_e. \quad (60)$$

The usual method of substitution for k' and integration by partial fractions gives

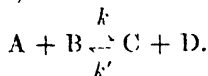
$$k = \frac{x_e}{t(a^2 - x_e^2)} \ln \frac{x_e(a^2 - x_e^2)}{a^2(x_e - x)}.$$

The isomerization of an alkyl ammonium cyanate to the corresponding substituted urea in aqueous solution, viz.,



is a process in which a second order reaction is opposed by one of the first order; an equation analogous to (61) has been found to be satisfactory (J. Walker and J. R. Appleyard, 1896).

The last example to be considered is that of a second order reaction opposed by one of the same order; thus,



If C and D are absent initially, whereas A and B are at the same concentration, then

$$\frac{dx}{dt} = k(a - x)^2 - k'x^2 \quad (62)$$

and at equilibrium

$$k(a - x_e)^2 = k'x_e^2, \quad (63)$$

in the usual notation. Substituting for k' and integrating, as in the previous cases, it is found that

$$k = \frac{x_e}{2ta(a - x_e)} \ln \frac{x(a - 2x_e) + ax_e}{a(x_e - x)}. \quad (64)$$

An interesting application of (64) is found in the reaction, investigated by M. Bodenstein (1899),



which is of the second order in both directions. If the initial concentration of hydrogen iodide is a , and x is the extent of decomposition after time t , the concentrations of hydrogen iodide, hydrogen and iodine are then $a - x$, $x/2$ and $x/2$ respectively, so that

$$\frac{dx}{dt} = k(a - x)^2 - k'(x/2)^2 \quad (65)$$

and at equilibrium

$$k(a - x_e)^2 = k'(x_e/2)^2. \quad (66)$$

If the value for k' is now substituted in (65) it is seen that the result is exactly the same as that obtained from (62) and (63), so that (64) should be applicable. Starting with pure hydrogen iodide gas and heating it to constant temperature until equilibrium was obtained, Bodenstein determined the value of x_e by analysis of the resulting system after rapid cooling (p. 831). The experiment was then repeated with a series of separate bulbs containing hydrogen iodide each of which was heated for a different length of time, and then suddenly chilled. The results obtained in one set of observations are given in Table 181. When the experimental

TABLE 181. DECOMPOSITION OF HYDROGEN IODIDE AT 410° C.

t	$(a = 4.46 \quad x_e = 0.938 \times 10^{-2} \text{ mole per liter})$	x	$k \text{ (liter mole}^{-1} \text{ sec.}^{-1})$
50 min.		2.79×10^{-2}	5.08×10^{-4}
100		4.95	5.04
150		6.16	4.59
200		7.37	4.89

difficulties are taken into consideration it is seen that the agreement between the values of k , obtained by (64), is satisfactory. The reverse second order reaction, between hydrogen and iodine molecules, has also

been studied and found to give results in accordance with the theoretical equation.⁵

It has been assumed in the above discussion that the reactants are at the same concentration and that the products are completely absent at the commencement. These restrictions do not necessarily apply in practice, and the general case for second order direct and reverse reactions would be

$$\frac{dx}{dt} = k(a - x)(b - x) - k'(c + x)(d + x), \quad (67)$$

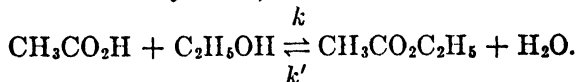
where a , b , c and d are the initial concentrations of the reactants A and B and the products C and D, respectively. Appropriate simplifications are made if C or D is absent, or if the reactions are not of the second order. The integration of such equations, after substituting the value of k' from the conditions of equilibrium, is complicated but it has been carried out and the results are to be found in the chemical literature.

Equilibrium Constant and Velocity Constants.—For any reversible reaction the equilibrium constant K is equal to the ratio of the velocity constants of direct and reverse reactions, provided the systems are virtually ideal so that the kinetic factors (p. 826) are almost unity; that is,

$$K = k/k'. \quad (68)$$

If k is found, therefore, by one of the appropriate velocity equations already given, then k' for the opposing reaction can be calculated if K is known. In any case the equations representing the condition of equilibrium, employed in the evaluation of k , give a relationship between k and k' which is virtually equivalent to determination of the equilibrium constant.

For some reactions it has been found possible to measure k and k' in the early stages when the extent of the opposing reaction is negligible in each case; the values obtained in this way, therefore, do not involve any reference to equilibrium. If the equilibrium constant is subsequently determined directly, the data provide a means of verifying the relationship between k , k' and K . For example, O. Knoblauch (1897) studied the rate of esterification of acetic acid by ethyl alcohol in the presence of water with hydrochloric acid as catalyst, and also the hydrolysis of the ester using the same catalyst viz.,



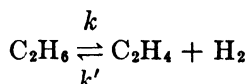
From the results obtained in the early stages of the two processes J. H. van't Hoff (1898) calculated the values of k and k' , the reverse reaction being ignored in each case; they were as follows:

$$k = 2.38 \times 10^{-4} \quad \text{and} \quad k' = 0.815 \times 10^{-4};$$

$$\therefore \frac{k}{k'} = K = \frac{2.38}{0.815} = 2.92.$$

The equilibrium constant calculated from the concentrations of reactants and products when equilibrium was attained, in the presence of the same amount of hydrochloric acid, was 2.84. The agreement is as good as could be expected when departure from ideal behavior is taken into consideration. It will be noted that the equilibrium constant recorded is appreciably lower than that given on p. 843; this is attributed to the difference in the medium and its effect on the activities.

The decomposition of ethane into ethylene and hydrogen, treated as a first order reaction, has been studied, and so also has the recombination process, considered as a simple second order process; hence k and k' for the reversible system



are known. Since measurements were made at different temperatures they may be compared by expressing the results in the general form

$$\log k = -\frac{73,170}{4.57T} + 15.12 \quad \text{and} \quad \log k' = -\frac{43,150}{4.57T} + 8.457,$$

with pressures in atmospheres; hence,

$$\log K_p = \log k - \log k' = -\frac{30,020}{4.57T} + 6.66.$$

By direct measurement at equilibrium it was found (R. N. Pease, 1928) that

$$\log K_p = -\frac{31,224}{4.57T} + 6.31$$

holds between 600° and 700° c. In view of the experimental difficulties and the necessity for extrapolation over a range of temperature, the agreement between the two equations for $\log K_p$ is good.

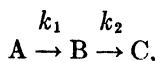
In order to obtain the results recorded above the velocity measurements were made on systems which were considerably removed from equilibrium conditions, and in verifying (68) it is assumed that the same specific reaction rates are applicable at equilibrium. A direct experimental proof of this view has been obtained by making use of artificially radioactive arsenic. A mixture of inactive arsenic acid, potassium iodide and active arsenious acid was made up, so that the equilibrium



was established in aqueous solution. The rate of reaction, at equilibrium, was then determined from measurements of the distribution of radioactivity between arsenate and arsenite after various intervals of time. It was demonstrated that there is no exchange in the absence of the iodide-iodine system, and so the formation of active arsenate must be due to the reaction under consideration.

The observed velocity was found to be in good agreement with that obtained normally with systems far removed from equilibrium (J. N. Wilson and R. G. Dickinson, 1937).

Consecutive Reactions.—The problem of consecutive reactions has been already considered in a general manner, the assumption being made that one of the stages was relatively much slower than the others, so that the latter did not appreciably influence the kinetics of the whole reaction. It is of interest to examine the subject of consecutive reactions when the different stages occur with speeds of similar magnitude. The simplest case arises when the consecutive processes are both of the first order, viz.,



where k_1 and k_2 are the two specific rates. If a is the concentration of the reactant A at the commencement, and c_A , c_B and c_C are the values for A, B, and C respectively, after time t , then

$$a = c_A + c_B + c_C. \quad (69)$$

The rate at which A disappears is given by

$$-\frac{dc_A}{dt} = k_1 c_A \quad \therefore c_A = ae^{-k_1 t} \quad (70)$$

[compare equation (5), p. 1046], and that at which C is formed from B is

$$\frac{dc_C}{dt} = k_2 c_B. \quad (71)$$

The rate of accumulation of B in the system is equal to the difference in its rate of formation from A by (70) and its decomposition to C by (71); hence,

$$\frac{dc_B}{dt} = -\frac{dc_A}{dt} - \frac{dc_C}{dt} = k_1 c_A - k_2 c_B, \quad (72)$$

and from (70), (71) and (72) it can be shown that

$$c_B = a \frac{k_1}{k_2 - k_1} (e^{-k_1 t} - e^{-k_2 t}). \quad (73)$$

Since c_A is given by (70) as $ae^{-k_1 t}$, the value of c_C can be readily obtained with the aid of (69). In order to test these deductions experimentally it is necessary to know k_1 and k_2 , but apart from the radioactive series, which represent a series of changes of the first order, there appear to be only a limited number of instances where the requisite data are available.

An interesting type of first order consecutive reactions occurs in the hydrolysis of ethyl succinate, ethyl tartrate and glycol diacetate in the presence of hydrochloric acid as catalyst. The reaction rate is followed by titration of the free acid, in excess of that used as catalyst, present after various intervals of

time (cf. p. 1053). In these instances k_1 is almost exactly equal to $2k_2$ and the result, as can be readily shown from the relationships developed above, is that the whole process will appear to be a simple reaction of the first order. The catalyzed hydrolysis of glyceryl triacetate (triacetin) shows a similar behavior because the specific rate constants of the three consecutive stages are in the ratio of 3:2:1 (J. Meyer, 1909).

A few reactions have been studied involving consecutive second order stages, but here the integration is so complicated that approximate methods have to be employed. Mention may be made of the saponification of ethyl succinate by sodium hydroxide which takes place in two second order steps, viz.,

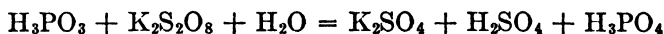


and



The velocity constant for the first reaction is obtained from the early stages of the hydrolysis of the ester, before appreciable amounts of the sodium ethyl succinate have formed; the second constant is derived from experiments on the saponification of the latter alone.

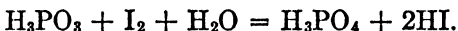
Another example is provided by the reaction



with hydrogen iodide as catalyst; the process occurs in two stages



and



These two reactions have been studied separately, and the kinetics of the whole process have also been examined; the results are in fair agreement with those expected from the approximate solutions of the appropriate differential equations.

Since consecutive first order reactions are more amenable to theoretical treatment, these will be discussed further. If k_1 and k_2 are known then it is possible, as shown above, to evaluate c_A , c_B and c_C , for a given initial concentration a , at different times after the commencement. The general nature of the results obtained is indicated by Fig. 246, which shows the concentrations of the three substances present after various times. The concentration of A decreases steadily from the beginning, since this decomposes in an exponential manner, irrespective of the fate of its product B. Similarly, the concentration of C must increase steadily, and ultimately the value must become equal to the initial concentration of A. On the other hand, the amount of the intermediate substance B increases at first, but as it accumulates the rate of decomposition becomes greater and so its concentration passes through a maximum, and finally it must, of course, become zero. By differentiation of (73) with respect to t , and

equating to zero, the conditions for the maximum are found to be

$$t_{\max.} = \frac{\ln k_1/k_2}{k_1 - k_2} \quad (74)$$

and

$$(C_B)_{\max.} = a \left(\frac{k_2}{k_1} \right)^{\frac{k_2}{k_1 - k_2}}. \quad (75)$$

If a system of two consecutive first order reactions is studied kinetically, it will be seen that a satisfactory first order constant may be ob-

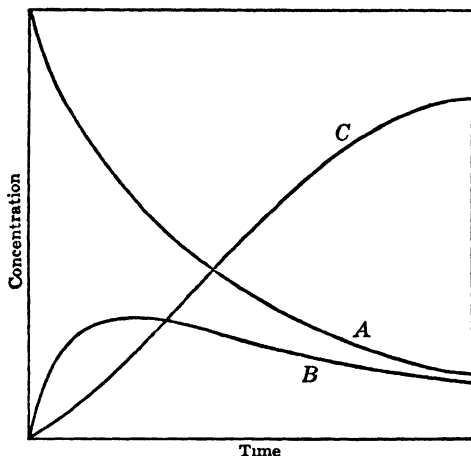
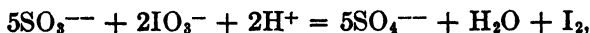


FIG. 246. Consecutive reactions $A \rightarrow B \rightarrow C$

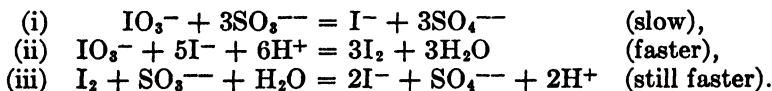
tained provided the amount of A remaining at various times, i.e., $a - x$, is determined *directly*, or if the amount decomposed x is obtained from the *sum* of B and C. When k_2 is large in comparison with k_1 it is apparent from (75) that the maximum concentration of B will be small, and so nearly the whole of the reaction product will be in the form of C; a reasonably good constant for the primary reaction could then be obtained by assuming x to be equal to the concentration of C at any instant. When different kinetic results are obtained for a given reaction according as $a - x$ is determined from the concentration of the reactant or from that of a product, it is a clear sign that the reaction is a complex one.

Period of Induction.—It has been found that certain reactions are characterized by the fact that there is an initial period during which the process appears to be quiescent. The existence of such a phenomenon was first realized by A. W. Cruickshank (1801), and subsequently rediscovered by J. Dalton (1811) and J. W. Draper (1843); to it has been given the name **period of induction** (R. Bunsen and H. E. Roscoe, 1855). There are probably different factors capable of producing these periods of induction, and that observed when a mixture of hydrogen and chlorine is

exposed to light has been traced to nitrogen compounds present as impurities (C. H. Burgess and D. L. Chapman, 1906); when these substances are removed the period of induction is no longer detected. Induction periods are observed in some process involving chain mechanisms (p. 1080) which, in a sense, may be regarded as consecutive reactions. An interesting illustration of induction is found in the familiar experiment of H. Landolt (1886), in which iodic and sulfurous acids react in the presence of starch; although the chemical process is



no blue color, due to the starch-iodine complex, is at first observed, but after some time, depending on the concentration of the solution, it suddenly appears. A study of the complex reaction, by J. Eggert (1917) and others, has shown that it takes place in at least three successive stages:



The iodide ions, gradually formed in reaction (i), are oxidized in (ii) to free iodine, but this is rapidly removed by process (iii); only when the concentration of the sulfite ion is reduced to such an extent that reaction (iii) becomes relatively slow, will there be sufficient free iodine in the solution to produce the blue color with starch.

Periods of induction are sometimes associated with autocatalysis; the reaction between permanganate and oxalic acid in the presence of sulfuric acid is initially slow, but soon becomes very rapid. This induction effect was found by A. V. Harcourt and W. Esson (1866) to result from the fact that manganous sulfate is a catalyst for the reaction; as the manganous ions accumulate so the oxidation of the oxalic acid becomes faster and faster. Similarly, the action of nitric acid on copper is at first very slow, but the nitrous acid produced acts catalytically, and so the rate of dissolution of the metal increases rapidly. Subsequently it falls off again, for the nitrous acid decomposes when its concentration exceeds a certain amount. The induction due to autocatalysis may be regarded as a result of successive reactions: in each case there is a slow stage followed by one or more rapid ones in which a product of the first stage is concerned.

Periodic Reactions.—Periodic phenomena are not uncommon in chemistry, but there do not appear to be many examples of chemical reactions taking place with a velocity which successively increases and decreases several times. It has been shown theoretically, however, that in certain circumstances such reactions are possible (A. J. Lotka, 1910, 1920; J. Hirniak, 1910); for example, if in the successive stages



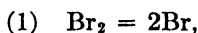
both reactions are autocatalytic and the amount of A is held constant, either by keeping the solution saturated, or by having a large excess of the

reactant, then a rhythmic increase and decrease in the rate of formation of C should be observed. It has been reported that, with suitable concentrations of the reactants, in the interaction of hydrogen peroxide and iodic acid, a process known to involve several successive reactions, the iodine is liberated in a periodic manner (W. C. Bray, 1921).⁶

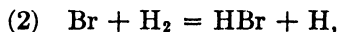
The Hydrogen-Bromine Reaction.—The thermal reaction between hydrogen and iodine, although reversible, is by no means complicated (p. 1072), but that between hydrogen and bromine vapor provides an important example of successive reactions. The process was studied at 200° to 300° c. by M. Bodenstein and S. C. Lind (1906) who concluded that the rate of formation of hydrogen bromide could not be expressed by any simple equation and proposed the complicated relationship

$$\frac{dc_{\text{HBr}}}{dt} = \frac{k c_{\text{H}_2} \sqrt{c_{\text{Br}_2}}}{m + \frac{c_{\text{HBr}}}{c_{\text{Br}_2}}} \quad (76)$$

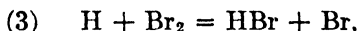
to represent the experimental results, where k and m are constants. The interpretation of this equation, developed by J. A. Christiansen (1919), K. F. Herzfeld (1919), and M. Polanyi (1920) is as follows. It is supposed that the first stage is the dissociation, to a small extent, of bromine molecules into atoms; thus,



which is followed by a relatively slow (endothermic) stage



and the rapid reactions



and



which do not necessarily occur in the order given. The specific rates are k_1, k_2, k_3, k_4, k_5 , respectively. Hydrogen bromide is produced in reactions (2) and (3), and destroyed in (4), and so its net rate of formation will be

$$\frac{dc_{\text{HBr}}}{dt} = k_2 c_{\text{Br}} c_{\text{H}_2} + k_3 c_{\text{H}} c_{\text{Br}_2} - k_4 c_{\text{H}} c_{\text{HBr}}. \quad (77)$$

To simplify this expression it is necessary to express the concentrations of hydrogen and of bromine atoms, c_{H} and c_{Br} , in terms of the corresponding molecules. This may be done by utilizing the general principle, based on theoretical and experimental considerations, that because the concentration of atoms (or radicals) is always relatively small, a so-called **stationary state** is reached soon after the reaction has started. The rate of change of the concentrations of atoms (or radicals) is so small, that these concentrations may be taken as constant. The rate of formation of atoms (or radicals) may thus be taken as equal to the rate at which

they disappear; for example, hydrogen atoms are produced by reaction (2) and removed by (3) and (4), and equating the two rates, it is seen that

$$k_2 c_{H_2} c_{Br} = k_3 c_{H} c_{Br_2} + k_4 c_{H} c_{HBr}. \quad (78)$$

Similarly, bromine atoms are formed in processes (1), (3) and (4) and destroyed in (2) and (5), so that in the stationary state

$$k_1 c_{Br_2} + k_3 c_{H} c_{Br_2} + k_4 c_{H} c_{HBr} = k_2 c_{H_2} c_{Br} + k_5 c_{Br}^2. \quad (79)$$

Subtracting (78) from (79) it follows that

$$\begin{aligned} k_1 c_{Br_2} &= k_5 c_{Br}^2; \\ \therefore c_{Br} &= \sqrt{k_1 c_{Br_2} / k_5}. \end{aligned} \quad (80)$$

Substituting this value for c_{Br} in (78), the concentration of hydrogen atoms is evaluated as

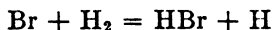
$$c_H = \frac{k_2 c_{H_2} \sqrt{k_1 c_{Br_2} / k_5}}{k_3 c_{Br_2} + k_4 c_{HBr}}. \quad (81)$$

If now these expressions for c_H and c_{Br} are inserted in (77) it is seen that

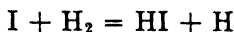
$$\frac{dc_{HBr}}{dt} = \frac{2k_2 \sqrt{k_1/k_5} c_{H_2} \sqrt{c_{Br_2}}}{1 + \frac{k_4}{k_3} \cdot \frac{c_{HBr}}{c_{Br_2}}}, \quad (82)$$

which is identical in form with the empirical equation (76). This is an important illustration of the application of the concept of successive reactions to the interpretation of processes which are kinetically complicated.⁷

Chain Reactions.—In the hydrogen-bromine reaction the stage



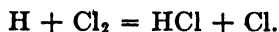
is endothermic to the extent of nearly 17 kcal., and so the process occurs relatively slowly (cf. p. 1092); the corresponding reaction involving iodine



is even more endothermic, viz., 35 kcal., and is so slow that it has no measurable influence on the reaction between hydrogen and iodine, although the iodine atoms are probably formed in greater concentration than are bromine atoms from the corresponding molecules. The reaction between chlorine atoms and molecular hydrogen, on the other hand, occurs so readily that the consequences are of great significance. If atomic chlorine is introduced in a suitable manner into a mixture of chlorine and hydrogen, generally by exposure to light (see p. 1173), the two following reactions apparently occur, both being very rapid:



and

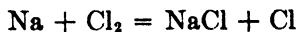


As a result a molecule of hydrogen and one of chlorine have been made to combine, but a chlorine atom has been regenerated and this can set off the combination of further hydrogen and chlorine molecules. The introduction of a few chlorine atoms can thus initiate a chain of consecutive reactions, the two stages repeating themselves until the reactants are used up. The series of successive processes is known as a **chain reaction**. Although it might at first appear that a chain has an unlimited length, there are factors operative tending to break the chain. In the hydrogen-chlorine reaction removal of hydrogen or chlorine atoms by any of the processes



will result in the termination of chains. These reactions are all, however, of the type requiring three-body collisions, so that they do not occur very readily in the gas phase, although they can take place at the walls of the vessel. At low pressures there is greater probability of the atoms reaching the walls, since the number of collisions in the gas phase is relatively small; in agreement with expectation it is found that the chain length, as determined from the yield of hydrogen chloride, is less at low than at higher pressures. The introduction of powdered glass or silica into the reaction vessel is often found to *decrease* the reaction velocity; if this occurs then it is reasonably certain that a chain reaction is in progress, for otherwise if increase of surface had any effect at all it would be to increase the velocity. In general, if an increase of the area/volume ratio of the reaction vessel decreases the velocity, a reaction is occurring in which chains are broken at the walls. In some chain reactions, however, e.g., the oxidation of benzene vapor, the majority of chains appear to be broken within the gas itself and not at the walls. As a rule the relative extents of the two methods of chain termination depend on the pressure (see p. 1083).

Although the chains in hydrogen-chlorine mixtures are usually initiated by chlorine atoms produced by light, this is not essential; the atoms may be the result of heating or better of introducing a small quantity of sodium vapor (M. Polanyi, 1927), when the reaction



occurs, and the chlorine atoms can set off the hydrogen-chlorine chain.

The concept of reaction chains was first postulated in connection with the photochemical combination of hydrogen and chlorine (p. 1173), but it has proved of great value in the interpretation of many other reactions, both thermal and photochemical. The chain carriers are usually atoms or radicals, and the resulting chains are referred to as **atomic chains** or **radical chains**. The possibility has been considered that chains may be

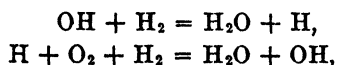
propagated by molecules of product carrying an excess of energy, obtained from the heat evolved in the reaction, which they transfer to molecules of reactant. Such **energy chains**, however, are apparently uncommon.

In addition to the effect of changing the area/volume ratio of the reaction vessel, the occurrence of atomic or radical chains can often be detected by the addition of small amounts of nitric oxide or propylene. These substances are apparently able to combine and remove the chain carriers, and so they produce a marked decrease in the reaction rate. The presence of atoms or radicals in a reacting system can frequently be proved by their acceleration of the conversion of para- to ortho-hydrogen, or *vice versa* (cf. p. 95). An important test for organic free radicals is based on the work of F. Paneth (1929 *et seq.*). By heating a tube through which lead tetramethyl (or tetraethyl) vapor is passed at low pressure, a mirror of metallic lead is formed; if a system containing free radicals, particularly methyl or ethyl radicals, is passed over the heated mirror, the latter will be removed. Mirrors of antimony, bismuth, tellurium, and of other elements which can combine with organic radicals, have been used for the same purpose. The deliberate introduction of atoms or free radicals, from another source, into a reacting system often provides useful information on the occurrence, or otherwise, of reaction chains.

A number of processes, as for example the oxidation of many gaseous hydrocarbons, carbon monoxide, hydrogen, hydrogen sulfide, carbon disulfide and methyl alcohol, the reaction between hydrogen bromide and ozone, and the thermal decomposition of ozone sensitized by bromine vapor, occur so rapidly that the formation of chains appears very probable. It is not always possible to state with any certainty the nature of the stages constituting the chains, but the general phenomena associated with the reactions show that chain mechanisms are operative. For example, the reaction between hydrogen and oxygen above 520° c. is retarded by the walls of the vessel and accelerated by inert gases. The rate of reaction is approximately proportional to the cube of the hydrogen concentration and to a power, greater than unity, of that of oxygen. This is evidently a chain reaction, the chains being broken at the walls; the molecules of inert gas prevent the carriers from reaching the walls and so increase the effective chain length. At high pressures increasing numbers of chains are, however, broken in the gas phase.

There is evidence to show that chains can originate not only in the interior of a gas but also at the walls of the vessel. This is perhaps somewhat surprising, since chains generally *terminate* at the walls, but it must be remembered that the two processes may be quite distinct. It was observed by H. N. Alyea and F. Haber (1930) that if streams of hydrogen and oxygen, at low pressure and about 530° c., meet at right angles to one another in the center of a large vessel, no reaction occurs, although if the gases are mixed in a smaller vessel of silica or porcelain under the same conditions an explosion results. If, however, a quartz rod is introduced at the place where the two streams of gas meet in the large vessel there is immediate reaction. It has been suggested that the hydrogen and oxygen first combine on the surface to form hydroxyl radicals which are

ejected into the gas and so initiate chains; thus,



the hydroxyl radical being destroyed and regenerated in the two successive reactions. Actually the combination of hydrogen and oxygen is a very complex reaction which is not yet fully understood, and many other simultaneous and consecutive reactions, in addition to those mentioned, undoubtedly take place. It is probable, too, that under different conditions the chains originate in different ways. In the oxidation of carbon disulfide vapor and of carbon monoxide, and in the reaction between ozone and hydrogen bromide, some, at least, of the chain carriers are formed at the walls of the vessel. In this way it is possible to account for the influence of the nature and area/volume ratio of the reaction vessel on reactions which are apparently homogeneous.

Kinetics of Chain Reactions.—The derivation of equations for the kinetics of successive reactions when a chain mechanism is operative is often difficult. An approximate equation, however, of the type

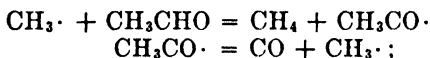
$$\text{Rate of reaction} = \frac{F(c)}{f_s + f_c + A(1 - a)} \quad (83)$$

appears to be applicable, where $F(c)$ is a function of the concentration of the reacting gases, f_s and f_c represent factors for the breaking of chains by a surface and by collisions in the gas, respectively, A is a constant many times larger than f_s and f_c , and a is the average number of molecules, or atoms, of chain carrier formed in each chemical act. In the cases considered above a has always been unity, so that the rate of reaction is finite and dependent on the chain length, the latter being related in an inverse manner to f_s and f_c ; such chains are said to be **stationary**. If, on the other hand, a is greater than unity, e.g., $\text{X} \rightarrow 2\text{Y}$, where X is the reactant and Y the chain carrier, then $1 - a$ is negative, and according to (83) the rate of reaction may become infinitely large. Each reacting molecule is replaced by more than one carrier, with the result that a number of branching chains are set up and the velocity increases with great rapidity until an explosion occurs; such chains are said to be **non-stationary**. The explosions accompanying them must be distinguished from ordinary thermal explosions which are due to the increase in temperature of the gas molecules by the heat generated in the initial stages of the reaction.

One of the results of nonstationary chains is the phenomenon of **explosion limits** discovered independently by N. Semenov (1927) and C. N. Hinshelwood (1928). In such processes as the combination of hydrogen, phosphine, carbon disulfide or carbon monoxide with oxygen, the rate of reaction increases steadily with increasing pressure, because of the accompanying increase of concentration. At a certain pressure, however, the reaction suddenly becomes explosive, its velocity being very great; over a range of pressure this explosive reaction persists, but

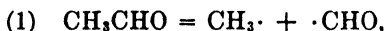
at another definite pressure the velocity decreases suddenly and the reaction behaves normally again. The simplest explanation of these pressure limits is that a nonstationary chain mechanism is operative, but at low pressures the carriers reach the walls of the vessel and are, to a great extent, destroyed. The rate of reaction is, therefore, not excessive. At a certain pressure, however, the rate of propagation of the branched chains exceeds their rate of destruction and the reaction becomes explosive; the sum of $f_s + f_c$ in (83) is then numerically less than $A(1 - a)$. Further increase of pressure favors the breaking of reaction chains by collisions in the interior of the gas, i.e., f_c is increased, and eventually a point is reached when $f_s + f_c$ is greater than $A(1 - a)$ and the speed is once more finite.

Free Radicals and Chain Reactions.—It appears to be established that free radicals play an important rôle in many homogeneous reactions and in several instances they are able to initiate reaction chains. For example, acetaldehyde vapor is stable at 300° c., but if a little azomethane is introduced the aldehyde immediately decomposes; the azo-compound yields methyl radicals, and it is probable that they start a series of chain reactions in which the acetaldehyde takes part. Although the actual reaction is undoubtedly complicated, as indicated below, the chain-carrying mechanism is probably

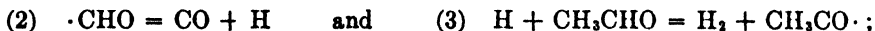


the methyl radical is regenerated and so is capable of carrying on the chain. The rate of reaction is markedly inhibited by packing the reaction vessel with glass tubing; this shows that chains are present which are terminated on surfaces.

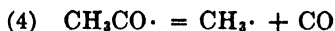
Not only is it possible for free radicals introduced in an extraneous manner to initiate reaction chains, but the radicals may be produced by the primary decomposition of the reactant itself, so that the first stage of the process will be followed by a chain of reactions (F. O. Rice and K. F. Herzfeld, 1934). As an illustration of the type of behavior which is possible, the following mechanism proposed for the thermal decomposition of acetaldehyde may be considered. The first stage is supposed to be the formation of $\text{CH}_3\cdot$ and $\cdot\text{CHO}$ radicals, thus



which is followed by the stages



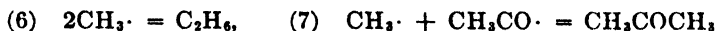
and the free radicals $\text{CH}_3\text{CO}\cdot$ and $\text{CH}_3\cdot$ then act as chain carriers by the reactions postulated above, viz.,



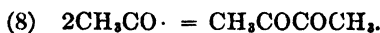
and



The removal of the carriers should then occur by the following processes:



and



In the stationary state (p. 1079) the rates of formation of the radicals $\text{CH}_3\cdot$, $\cdot\text{CHO}$, H and $\text{CH}_3\text{CO}\cdot$ are equal to their respective rates of removal. If the concentrations of CH_3CHO , $\text{CH}_3\cdot$, $\cdot\text{CHO}$, H and $\text{CH}_3\text{CO}\cdot$ in the stationary state are x_1 , x_2 , x_3 , x_4 and x_5 , respectively, and the specific rate constants of the eight reactions postulated above are k_1 , k_2 , k_3 , k_4 , k_5 , k_6 , k_7 and k_8 , respectively, then since the rates of removal and formation are equal for both $\cdot\text{CHO}$ and H , it follows that

$$k_1x_1 = k_2x_3 \quad \text{and} \quad k_2x_3 = k_3x_4x_1; \\ \therefore x_4 = k_1/k_3. \quad (84)$$

Similarly, for $\text{CH}_3\cdot$, and $\text{CH}_3\text{CO}\cdot$,

$$k_1x_1 + k_4x_5 = k_5x_2x_1 + k_6x_2^2 + k_7x_2x_5 \quad (85)$$

and

$$k_3x_4x_1 + k_5x_2x_1 = k_4x_5 + k_7x_2x_5 + k_8x_5^2. \quad (85a)$$

Because of the heats of activation (p. 1091), it is probable that the $\text{CH}_3\cdot$ radicals will be the most abundant chain carriers; it is thus permissible to neglect the terms $k_7x_2x_5$ and $k_8x_5^2$ in (85a). It follows, therefore, from (84), (85) and (85a) that

$$x_2 = (2k_1x_1/k_6)^{\frac{1}{2}}.$$

The rate of disappearance of acetaldehyde, by stages (1), (3) and (5), gives the over-all rate of the reaction; thus,

$$-\frac{dx_1}{dt} = k_1x_1 + k_3x_4x_1 + k_5x_2x_1,$$

and insertion of the values for x_4 and x_2 , gives

$$-\frac{dx_1}{dt} = 2k_1x_1 + k_5(2k_1/k_6)^{\frac{1}{2}}x_1^{\frac{3}{2}}.$$

If stage (1) is relatively slow, as it may well be, this rate equation reduces to

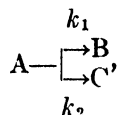
$$-\frac{dx_1}{dt} = k'x_1^{\frac{3}{2}}, \quad (86)$$

so that the chain reaction should have an apparent order of 1.5; this has been actually observed for the decomposition of acetaldehyde under certain conditions. It is of interest to note that if (7) had been the chief chain-breaking reaction, the over-all process would have been roughly of the first order, in spite of its complexity. These deductions show that care must be taken in interpreting the results of reaction kinetics if chains are involved; the matter will be considered further in connection with the theory of unimolecular reactions.

Sufficient has been said to indicate the importance of chain reactions in chemical kinetics, and further reference will be made to them in the discussion of homogeneous catalysis, negative catalysis in solution, photochemical processes, and other topics.⁸

Simultaneous Side Reactions.—In some processes the reacting substance can be removed by two or more reactions occurring simultaneously

and side by side; this may be depicted as



where A is the reactant. If a is the initial concentration of A, and x is the amount decomposed after time t , then

$$\frac{dx}{dt} = k_1(a - x) + k_2(a - x) = (k_1 + k_2)(a - x), \quad (87)$$

where k_1 and k_2 are the specific rates of the two side reactions, both being assumed to be of the first order. By integration it follows directly that

$$k_1 + k_2 = \frac{1}{t} \ln \frac{a}{a - x}, \quad (88)$$

so that the reaction as a whole behaves as if it were a simple one of the first order, the observed value of k being the sum of the specific rates of the two simultaneous reactions. In order to determine k_1 and k_2 separately, use is made of the following facts:

$$\begin{aligned} \text{Rate of formation of B} &= k_1(a - x), \\ \text{Rate of formation of C} &= k_2(a - x), \end{aligned}$$

and hence at any time the ratio of the concentrations of B to C is equal to k_1/k_2 . If these concentrations are determined at any stage, e.g., at the end of the reaction, the ratio of the velocity constants is at once known.

The fact that the ratio of the concentrations of B to C at any instant is a constant, k_1/k_2 , and independent of time, is the basis of the method proposed by R. Wegscheider (1899) as a test for side reactions. The action of hydrochloric acid on cinchonine yields two products, viz., a hydrochloride of cinchonine and a hydrochloride of an isomeric form of cinchonine; Z. H. Skraup (1899) thought there were two consecutive reactions and that the former product represented an intermediate stage in the formation of the latter. This view was shown to be incorrect by the observation that the relative amounts of the two hydrochlorides remained constant and did not vary with time, so that they were produced by two independent reactions occurring side by side. It will be evident from what has been stated on p. 1075 that it is very improbable that such a result would have been obtained if there were two consecutive reactions. The work of A. F. Holleman (1899) on the nitration of various benzene derivatives showed that the proportions of the *ortho*-, *meta*-, and *para*-nitro-compounds formed were the same throughout the reaction, indicating that they were produced in three side reactions. In alcoholic solution

camphorcarboxylic acid decomposes, giving off carbon dioxide, and at the same time esterification takes place; the two processes are independent side reactions as shown by the amounts of the products formed after various time intervals (G. Bredig and R. W. Balcom, 1908).

It can be readily shown that the Wegscheider test for side reactions is applicable only if both changes are of the same order; if the orders are different, the ratio of the concentrations of the products will vary over a period of time. Consider, for example, the possibility that

$$\begin{aligned}\text{Rate of formation of B} &= k_1(a - x)^{n_1}, \\ \text{Rate of formation of C} &= k_2(a - x)^{n_2};\end{aligned}$$

it is evident that if n_1 and n_2 are not equal the ratio of the amounts of B and C formed will change as the reaction proceeds.

A special case of side reactions is frequently encountered in acid-base catalysis (p. 1132); for example, in the mutarotation of glucose a series of pseudo-unimolecular side reactions occur, each involving a molecule of glucose and one or other of a number of catalysts, the concentrations of which remain constant. Thus, if c_G represents the concentration of α -glucose, then

$$-\frac{dc_G}{dt} = k_1c_1c_G + k_2c_2c_G + k_3c_3c_G + \cdots, \quad (89)$$

$$= kc_G, \quad (89a)$$

where k is equal to $k_1c_1 + k_2c_2 + k_3c_3 + \cdots$, c_1, c_2, c_3 , etc., representing the constant concentrations of various catalytically active species, and k_1, k_2, k_3 , etc., are the corresponding velocity constants. The same kind of equations apply to other catalytic processes in solution, such as the enolization of acetone, which is the rate determining stage in the reaction with iodine or bromine, the hydrolysis of esters, etc. (see p. 1134).⁹

THEORY OF REACTION RATES

The Temperature Coefficient.—Increase of temperature almost invariably increases the velocity of a chemical reaction to a marked extent, and for homogeneous processes the specific rate is approximately doubled or trebled for each 10° rise of temperature; this is sometimes expressed in the form of the **temperature coefficient**, $k_{t+10}/k_t \approx 2$ to 3, where k_t is the specific rate at t° and k_{t+10} at 10° higher. It is only an approximate method for indicating the effect of temperature because the coefficient decreases with increasing temperature. The results in Table 182 show

TABLE 182. INFLUENCE OF TEMPERATURE ON CHEMICAL REACTION VELOCITY

I		II	
Temperature	k	Temperature	k
15° c.	9.67×10^{-6}	300° c.	2.91×10^{-6}
60°	6.54×10^{-4}	400°	8.38×10^{-4}
101°	3.18×10^{-2}	500°	7.65×10^{-2}

the effect of temperature on the velocities of two entirely different reactions: (I) the decomposition of dibromosuccinic acid in aqueous solution (L. C. Schwab, 1883), and (II) the dissociation of gaseous hydrogen iodide (M. Bodenstein, 1899). The temperature coefficient for the reaction I, about 1.9, is somewhat greater than that for II, about 1.7, but reactions with much higher coefficients are known; for example, k_{t+10}/k_t for the combination of methyl iodide and sodium ethoxide in solution is said to be 2.9, and for the decomposition of nitrogen pentoxide it is approximately 3.8 at room temperatures. The marked influence of temperature is somewhat surprising on first consideration, since the number of collisions between molecules (cf. p. 277) will increase only by about 2 per cent for a rise of 10° at ordinary temperature, whereas the reaction velocity may increase 100 or 200 per cent, or more. The explanation of this increase forms an important aspect of the study of reaction kinetics.

The Arrhenius Equation.—The most satisfactory method for expressing the influence of temperature on reaction velocity is that used by S. Arrhenius (1889), following J. H. van't Hoff (1884). The variation of equilibrium constant with temperature is given by the van't Hoff equation (p. 829); thus,

$$\frac{d \ln K_e}{dT} = \frac{\Delta E}{RT^2}, \quad (90)$$

and since K_e is equal to k/k' (p. 1073),* it follows that

$$\frac{d \ln k}{dT} - \frac{d \ln k'}{dT} = \frac{\Delta E}{RT^2}. \quad (91)$$

It was suggested by van't Hoff that (91) may be split up to give

$$\frac{d \ln k}{dT} = \frac{E}{RT^2} + A \quad \text{and} \quad \frac{d \ln k'}{dT} = \frac{E'}{RT^2} + A, \quad (92)$$

where $E - E'$ is equal to ΔE and A is a constant; it was found experimentally that the temperature variation of the specific rate was best expressed by assuming A to be zero, so that

$$(a) \quad \frac{d \ln k}{dT} = \frac{E}{RT^2} \quad \text{or} \quad (b) \quad \ln k = -\frac{E}{RT} + \text{const.} \quad (93)$$

These are (a) the Arrhenius equation and (b) its integrated form; they require the plot of $\ln k$, or $\log k$, against $1/T$ to be a straight line, a relationship which has been found to hold for a large number of reactions. The points in Fig. 247, for example, are plotted from the results of M. Bodenstein (1899) on the decomposition of hydrogen iodide; the linear relationship between $\log k$ and $1/T$ is evident. The Arrhenius equation is widely applicable not only to homogeneous gas reactions, but also to

* It is assumed that the specific reaction rates k and k' are expressed in terms of concentrations.

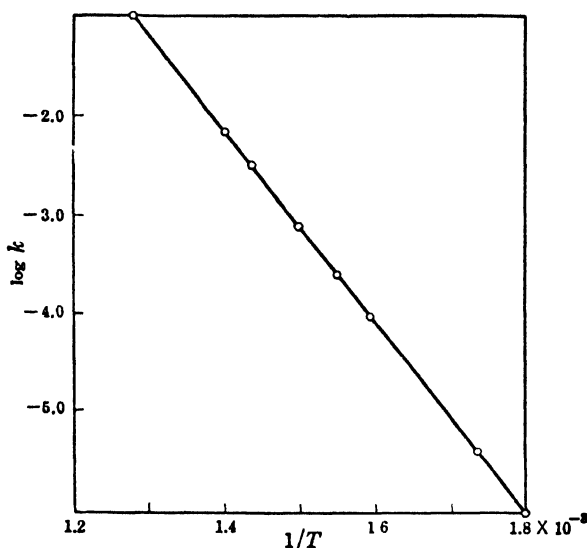


FIG. 247. Decomposition of hydrogen iodide (Bodenstein)

reactions in solution and to heterogeneous processes. It frequently fails, however, for chain reactions.

The energy E , called the "energy of activation," is of great theoretical importance, as will be seen shortly. It is determined from measurements of the specific rate at two or more temperatures; the value of E can then be obtained directly from (93), or by plotting $\ln k$ against $1/T$, the slope of the straight line (cf. Fig. 247) being equal to $-E/R$.

Molecular Activation.—To explain (93), Arrhenius suggested that in every system an equilibrium existed between normal and "active" molecules, and that only the latter could take part in chemical reaction. If heat is absorbed in the conversion of a normal to an active molecule, the van't Hoff equation (90), p. 1088, immediately gives a relationship like (93), where E represents the energy difference between an active and a normal molecule. This, then, accounts for the marked influence of temperature on reaction velocity, for a rise of temperature will favor the formation of active molecules, in accordance with the principle of mobile equilibrium. Although the number of collisions between molecules is affected only slightly, the concentration of active molecules may be doubled by a 10° rise of temperature, so that the velocity of reaction is greatly increased.

The suggestion that only active or "activated" molecules, with energy in excess of the average, can take part in reactions is the basis of modern views on the mechanism of chemical change. At one time it was thought that a molecule became activated as the result of an absorption of radiation, and this theory, known as the "radiation hypothesis" of chemical

reaction, received support from many workers, among whom may be mentioned M. Trautz (1909), J. Perrin (1916) and W. C. McC. Lewis (1916); various experiments, however, have shown this view to be unsatisfactory. Many substances have no absorption bands in the wave length region corresponding to that expected for the activating radiation, and exposure to radiations of various wave lengths in the infrared, the theoretically active region, has produced no increase in reaction velocity. The necessity for molecular collisions, even in a unimolecular reaction, was proved by the passage of a low-pressure "molecular beam" of pinene, similar to the beam employed in the Stern-Gerlach experiment (p. 64), through a tube in which it was subjected to thermal radiation. If the mean free path of the molecules was such that no collisions occurred in the tube, then no racemization reaction was observed. Radiation alone is clearly insufficient for chemical reaction (G. N. Lewis and J. E. Mayer, 1927).¹⁰

At the present time it is generally accepted that molecules acquire the additional energy required for reaction, known as the **energy of activation**, as a result of interchanges occurring in collisions. As seen in connection with the discussion of the distribution of molecular velocities and energies in Chapter IV, the molecules of a gas do not all have the same energy; a small fraction will have energies greatly in excess of the average. It is only molecules possessing energy in excess of a certain amount, viz., the energy of activation, which are capable of taking part in reaction. The magnitude of this energy will depend on the nature of the process, and so the proportion of "activated" molecules will vary from one reaction to another.

The probability that a molecule will possess energy in excess of an amount E per mole, at the temperature T , is related to the familiar Boltzmann factor $e^{-E/RT}$, where R is the molar gas constant. If the energy is restricted to translational energy in two components for a single molecule, or in one component for each of two molecules, making a total of two "square terms," then the fraction of molecules having energy in excess of E is actually equal to $e^{-E/RT}$ [cf. equation (86), p. 270]. For simplicity this will be taken as giving the condition that a pair of molecules have the energy necessary for reaction. If A , which may be called the **frequency factor**, represents the total frequency of encounters between two reactant molecules, irrespective of whether they possess sufficient energy or not, the reaction rate will clearly be dependent on the product of A and $e^{-E/RT}$. By appropriate choice of units of the frequency factor, this product may be identified with the specific reaction rate, so that it is possible to write

$$k = Ae^{-E/RT}. \quad (94)$$

Upon taking logarithms the result is

$$\ln k = -\frac{E}{RT} + \text{const.},$$

assuming A to be constant and almost independent of temperature for a given reaction. This is identical with the empirical equation (93) representing the experimental variation of specific rate with temperature, and hence it follows that the specific rate itself must be represented by an expression of the form of (94). It is of interest to show that the exponential factor accounts for the marked influence of temperature on the rate of a chemical reaction. For example, if the activation energy E is 24,000 cal. per mole, which is a reasonable value for a process taking place at ordinary temperatures, the value of $e^{-E/RT}$ at 300°K. is $e^{-24,000/2 \times 300}$, i.e., 4.25×10^{-18} , while at 400°K. it is $e^{-24,000/2 \times 400}$, i.e., 9.37×10^{-14} . An increase of 100° in the temperature is thus seen to increase the factor about 2×10^4 fold; this is approximately the same as that observed for reaction I in Table 182. The significance of this large increase in $e^{-E/RT}$ as the temperature is raised is that the proportion of molecules having energy in excess of the activation value E increases greatly with temperature; this is in agreement with the conclusion reached earlier (p. 1089).

It is evident that the larger the value of E the smaller will be the factor $e^{-E/RT}$, and since for many reactions the frequency factor does not vary greatly, it is the energy of activation which often determines whether a reaction is fast or slow at a given temperature. A bimolecular reaction having a value of E of about 20,000 cal. per mole will have a measurable velocity at about ordinary temperatures, but if E is as high as 40,000 cal. per mole, the velocity only becomes convenient for study in the region of 400°C. Thus, the larger the activation energy, the higher the temperature, in general, at which the reaction rate becomes appreciable. It should be noted that the occurrence of reaction chains of considerable length would, of course, vitiate these conclusions.

The Activated Complex.—It is evident from (94) that a complete theory of reaction rates involves an interpretation of the two quantities A and E , that is, the frequency factor and the energy of activation. The latter will be considered first. It is believed that when two reactant molecules, possessing the necessary energy of activation, come together, they first form an **activated complex** or **transition state**, and this decomposes at a definite rate to yield the products of the reaction. The formation of this intermediate state is considered to be characteristic of all chemical changes, and even of certain physical processes taking place at a definite rate. Consider, for example, the reaction



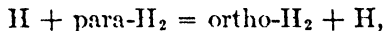
the atom A being gradually brought up to the diatomic molecule BC , in which the atoms are vibrating at their normal distance apart. As long as the reactants are relatively far from each other the potential energy of the system is unaffected, but as A approaches BC more closely, the nuclei of the latter are forced apart somewhat and the potential energy increases. This continues until a configuration, which may be represented as $A-B-C$

of a system of three atoms is given by

$$E = A + B + C + [\frac{1}{2}\{(\alpha - \beta)^2 + (\alpha - \gamma)^2 + (\beta - \gamma)^2\}]^{\frac{1}{2}}, \quad (95)$$

where $A + \alpha$, $B + \beta$ and $C + \gamma$ are the potential energies of the three possible pairs AB, BC and CA, respectively, in the absence of the third atom. The potential energy terms are divided into two parts, viz., A , B and C , and α , β and γ , respectively; the former are the Coulomb, or electrostatic, energies and the latter the exchange, or resonance, energies (p. 12). The sum of the two kinds of energy for various distances between the two atoms in the molecule can be evaluated from the appropriate Morse equations derived from spectroscopic data (p. 573), as shown by H. Eyring and M. Polanyi (1931). This is, however, not sufficient, for in order to calculate E by (95) it is necessary to know the separate values of the Coulomb and exchange energies; hence the assumption is generally made that the former constitutes about 10 to 20 per cent of the total binding energy. This assumption can be justified by the knowledge that in the hydrogen molecule, for which detailed calculations have been made (p. 102), the exchange energy constitutes from 10 to 15 per cent of the total potential energy at reasonable interatomic distances.

If the data are available for the three Morse equations to be determined, then with the aid of the postulate concerning the relative proportions of the Coulomb and exchange energies, it is possible to calculate E by (95) for various positions of the atoms A, B and C. For purposes of calculation the simplifying assumption is generally made that the three atoms are in a straight line, since this is the configuration of lowest potential energy and hence represents the path which the majority of molecules are likely to take. The results are best plotted in the form of a contour diagram, with the distances A to B and B to C as ordinates and abscissae, respectively, points of equal potential energy being joined by a series of contour lines. The potential energy diagram for the reaction



that is, the conversion of para-hydrogen to the ortho-form by means of atomic hydrogen, is shown in Fig. 249; * the calculations are based on the assumption that 14 per cent of the binding energy between pairs of atoms is Coulombic in nature. The diagram shows clearly a valley of minimum energy, indicating the approach of a hydrogen atom to a normal molecule of ortho-hydrogen; the level rises as the atom approaches and the energy reaches a maximum at a "saddle point," for the activated complex. From here the system can either return the way it came, or descend into the other energy valley representing the state of the products. The curve in Fig. 248 is, in fact, the energy path from one valley to the other drawn continuously in one plane, and the difference in energy between the valley representing $\text{H} + \text{H}_2$ and the top of the

* The angle of inclination of the axes is deliberately chosen so that a frictionless ball sliding on the potential energy surface would have the same distribution of translational and vibrational energy as the system under consideration.

"pass," i.e., the activated complex, is the activation energy of the reaction which can thus be evaluated. The energy at the highest point over which the reacting system has to pass is about 14 kcal., assuming 14 per cent Coulomb energy; if the latter is taken as 20 per cent of the total binding energy, the energy at the top of the pass is found to be 7.6 kcal. The experimental value for the activation energy of the reaction under consideration is approximately 6 kcal.

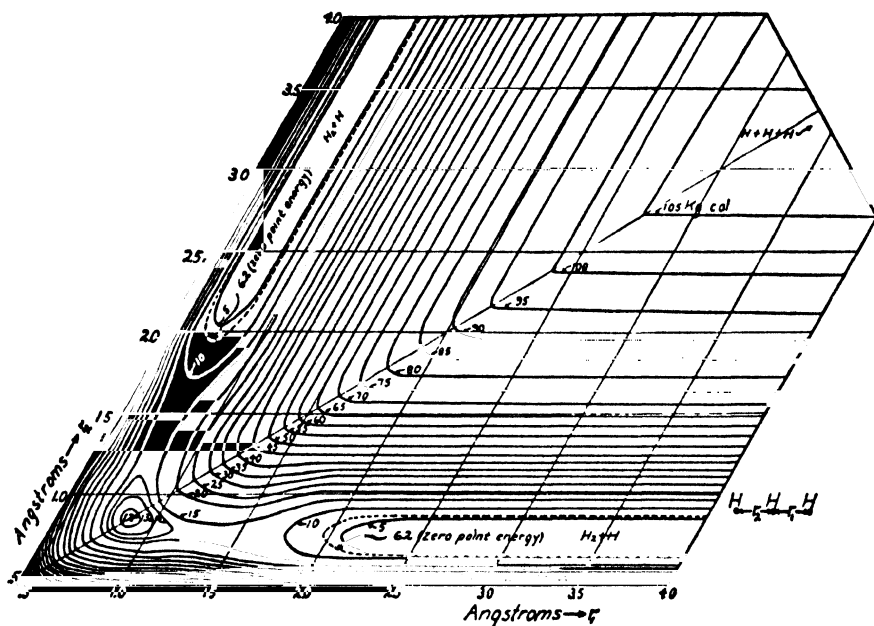


FIG. 249. Potential energy surface for $\text{H} + \text{H}_2 = \text{H}_2 + \text{H}$

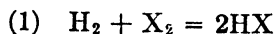
Working in a manner somewhat similar to that just described it is also possible to determine the potential energies for various distances apart of four atoms, so that the activation energy of a reaction of the type



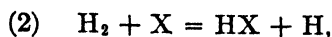
can be calculated.

The foregoing procedure has been applied to determine the energies of activation of a number of simple reactions and the results have often been found to be in satisfactory agreement with the experimental values. However, the approximations involved in the calculations, e.g., in the derivation of (95) and in the arbitrary choice of the proportion of Coulombic energy, make the quantitative conclusions somewhat uncertain. Nevertheless, a number of interesting qualitative results have been obtained, as the following examples show. Reaction between hydrogen and

halogen can take place either through atoms or molecules of the latter, viz.,



or



where X is the halogen. For the hydrogen-iodine reaction the calculations show that (1) should have the lower activation energy, whereas with bromine it may be lower for process (2), after making allowance for the energy required to form the halogen atom from the molecule. That is to say, in the reaction between hydrogen and iodine combination with molecules of halogen should occur more readily than with atoms, but for the hydrogen-bromine reaction the reverse may hold. Further, it appears that the homogeneous reaction between gaseous hydrogen and fluorine should require a relatively high energy of activation and hence should be a slow process (H. Eyring, 1931). This conclusion is somewhat surprising at first sight, but it is in agreement with experiment.¹¹

Simple Collision Theory of Reaction Rates.—In some of the earlier theories of reaction rates the frequency factor A in (94) was related directly to the number of collisions between the molecules, as derived from the simple kinetic theory of gases (W. C. McC. Lewis, 1918; C. N. Hinshelwood, 1923). The arguments used were as follows, for a bimolecular reaction involving two molecules of the same kind, e.g., the dissociation of hydrogen iodide (p. 1072). If z is the number of molecules in 1 ml. of gas taking part in binary collisions per second, and E is the energy of activation of the process, that is the minimum energy which the colliding molecules must possess before they can react, the number reacting is $ze^{-E/RT}$ per ml. per sec. It follows, therefore, that

$$\frac{dx}{dt} = ze^{-E/RT}. \quad (96)$$

Since the reaction involves two molecules of the same kind, the relationship between reaction velocity and concentration is

$$\frac{dx}{dt} = kn^2, \quad (97)$$

where k is the specific rate in appropriate units, and n is the concentration of reactant expressed as the number of molecules per ml.; from (96) and (97), therefore,

$$k = \frac{z}{n^2} e^{-E/RT}. \quad (98)$$

The value of z is known from the kinetic theory; it is *twice* the number of collisions per ml. per sec., since two molecules are concerned in each collision; hence (see p. 277),

$$z = 4n^2\sigma^2(\pi RT/M)^{\frac{1}{2}}, \quad (99)$$

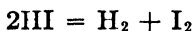
where σ is the molecular collision diameter and M the gram molecular weight. Substitution in (98) gives

$$k = 4\sigma^2(\pi RT/M)^{\frac{1}{2}}e^{-E/RT} \quad (100)$$

$$= Ze^{-E/RT}, \quad (101)$$

the quantity $Z = 4\sigma^2(\pi RT/M)^{\frac{1}{2}}$, which is equivalent to the frequency factor A in (94), being called the **collision number**; it is the number of collisions per second when there is only one molecule of reactant per ml.

Provided the collision diameter σ is known, the collision number can be calculated, and it is possible to compare it with the experimental value of the frequency factor. One way of performing this comparison is to calculate the specific rate of a reaction, whose activation energy is known, by utilizing (100) and then to see how this agrees with the observed rate. As an example of the calculation, the reaction



may be considered. From the work of Bodenstein (see Fig. 247) the energy of activation is found to be 44,000 cal. per mole, and at 556° K., assuming the collision diameter of the hydrogen iodide to be 3.5×10^{-8} cm., (99) gives for the gas at a concentration of 1 mole per liter, i.e., 6.02×10^{20} molecules of reactant per ml.,

$$6.0 \times 10^{31} \times e^{-44,000/2 \times 556} = 3.25 \times 10^{14}$$

as the number of molecules reacting per ml. per sec.; to convert this into moles per liter, it is necessary to multiply by 1000 and divide by the Avogadro number 6.02×10^{23} . Hence the number of moles of hydrogen iodide decomposed per liter per sec. when the concentration is 1 mole per liter is

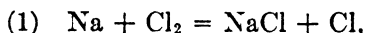
$$k = \frac{3.25 \times 10^{14}}{6.02 \times 10^{20}} = 5.4 \times 10^{-7},$$

which is the value of k in liter mole⁻¹ sec.⁻¹ units. The experimental specific rate at 556° K. is 3.5×10^{-7} . The agreement between the two results is good; in fact a deviation between calculated and observed values within a factor of 10 would be regarded as moderately satisfactory.

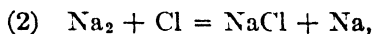
When the bimolecular reaction involves two different molecular species, the collision number Z , assuming one molecule of each kind per ml., is equal to $\sigma^2[8\pi RT(M_1 + M_2)/M_1M_2]^{\frac{1}{2}}$ where σ is the *mean* of the molecular collision diameters of the two reactants, and M_1 and M_2 are their molecular weights (p. 276). As before, the equation $k = Ze^{-E/RT}$ gives the specific rate with concentrations in molecules per ml., and the time in seconds. The energy of activation of the bimolecular reaction between hydrogen and iodine, also studied by M. Bodenstein (1899), has been employed to calculate the specific rate; it is found to be 14×10^{-2} at 700° K., in comparison with the observed value 6.4×10^{-2} liter mole⁻¹ sec.⁻¹. Here again the agreement is good.

The equations developed above will, of course, be applicable only for a simple homogeneous reaction, and will not apply if a chain mechanism is operative or if the process is heterogeneous. In addition to the examples mentioned above there are relatively few cases whereby the equations can be tested; the decompositions of nitrous oxide, chlorine monoxide, nitryl chloride (NO_2Cl), and gaseous acetaldehyde are not simple, but making some allowance for the complexities it appears that the calculated specific rate is at least of the correct order of magnitude. Satisfactory agreement of observed and calculated reaction rates have also been obtained for the decomposition of nitrosyl chloride and formaldehyde, for the reactions between ethylene and hydrogen and ethylene iodide and iodine atoms, and for a number of atom-molecule reactions involving hydrogen and deuterium (cf. p. 1060).¹²

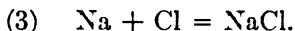
In certain reactions involving atoms the energy of activation appears to be zero, and every collision results in combination; such processes have been observed by M. Polanyi and his collaborators (1928 *et seq.*) in reactions between sodium vapor and a halogen or halogen compound. For example, if sodium vapor, containing Na and Na_2 molecules, is mixed with chlorine at low pressure, the following reactions occur:



followed by



and



The energy produced in reaction (2) is sufficient to excite a sodium atom* which, in returning to its normal state, emits the familiar *D*-lines; the process is, therefore, chemiluminescent (cf. p. 1188), and the light produced has been called a "highly dilute flame" or "diffusion flame," because of the very low pressure involved and the method of carrying out the experiment. Calculations show that reactions (1) and (2) occur at every collision, and no activation energy is required. Reaction (3), however, is of the type requiring a three-body collision (p. 1064), since *two* atoms are involved, and only occurs once in about 10,000 collisions. Other reactions similar to (1) and (2) have been found to require no energy of activation, although it must be pointed out that not all processes in which only one atom takes part are of this type. The reactions between sodium atoms and alkyl or aryl halides, for example, require activation energies of the order of 5,000 to 10,000 cal. per mole (H. von Hartel and M. Polanyi, 1932).¹³

Failure of Simple Collision Theory.—Although the simple collision theory of reaction rates, as described in the preceding section, is evidently satisfactory when applied to a number of reactions, it fails conspicuously in many cases. Apart from certain very rapid processes, which undoubtedly involve a chain mechanism, there are numerous instances in which

* Apparently not the one produced in the reaction, as explained later (p. 1189).

the reaction rate is much less than is to be expected from theoretical considerations. The bimolecular polymerization of ethylene, for example, occurs only once in about 2,000 collisions between *activated* molecules, and in the polymerization of 1:3-butadiene reaction takes place at one in 10,000 activated collisions. The combination of triethylamine and ethyl iodide in the gas phase at 140° c. occurs at a rate which is slower by a factor of about 10^{-8} than the calculated velocity,* and discrepancies of the order of 10^{-5} to 10^{-6} have been found for the reactions between ethyl alcohol and acetic anhydride vapor at 79° c., and between carbonyl sulfide and water vapor at about 300° c. A sufficient number of "slow" reactions are known to show that the phenomenon is not limited to a few instances, and (101) is modified empirically so as to take the form

$$k = PZe^{-E/RT}, \quad (102)$$

where P , the probability factor, or steric factor, makes allowance for any effects causing deviations from the "ideal" behavior of (101). The value of P varies from unity † to about 10^{-8} according to the nature of the reaction.¹⁴

Theory of Absolute Reaction Rates.—An alternative point of view, developed by M. Polanyi (1935), and particularly by H. Eyring (1935), following the earlier work of R. C. Tolman (1927) and H. Pelzer and E. Wigner (1932), provides a much more complete interpretation of the significance of the frequency factor in the reaction rate equation. It shows why the simple collision theory equation (100) is satisfactory in certain instances, and why it fails in the cases of the type mentioned above. Because the complete theory provides, in principle, a method for calculating the rate of a chemical reaction, using only such fundamental physical properties as the dimensions, vibration frequencies, masses, etc., of the reacting molecules, the treatment is referred to as the **theory of absolute reaction rates**. Before two (or more) molecules, possessing the requisite energy, can react they must first collide and form the activated complex for the reaction which then decomposes; the problem is to calculate the frequency with which this happens.

An examination of the potential energy surface in Fig. 249 shows that in all directions except that leading into the valleys, which may be called the reaction coordinate, any change in the dimensions of the activated complex leads to an increase of potential energy, ‡ just as is the case with an ordinary molecule (cf. Fig. 126, p. 573). It appears, therefore, that

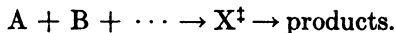
* The reaction that does occur appears to take place mainly on the surface of the vessel; the homogeneous gas reaction is thus probably even slower.

† Values of P as large as 10^6 have been observed for reactions between oppositely charged ions in solution.

‡ The surface in Fig. 249 shows a small shallow basin, about 1.5 kcal. deep, at the top of the energy pass; if this has any significance, it implies some stability of the molecule H, with respect to the activated state. The matter is not important, however, for the present discussion of reaction rates.

the activated complex may be treated by statistical methods just like a normal molecule, except that in addition to having three translational degrees of freedom, it has a fourth degree of freedom of movement along the reaction coordinate. It is in the direction of this coordinate that the activated complex approaches the top of the energy barrier, crosses it and then falls to pieces. Since a molecule can have only $3n$ energy variables in all (p. 331), if there are n atoms in the molecule, and since the activated complex is assumed to have an additional degree of translational freedom, it must have one vibrational degree of freedom less than a normal molecule with the same number of atoms. In other words, the activated complex may be treated as a molecule in which one of the vibrations is very stiff, so that the particular frequency is extremely large (H. Eyring, 1935).

Consider a process involving the reacting substances A, B, etc., which form the activated complex X^\ddagger as an intermediate stage in the reaction



The rate of the reaction is then equal to the concentration of the activated complexes at the top of the energy barrier multiplied by the frequency of crossing the barrier. If a length δ in the reaction coordinate, at the top of the barrier, is taken as representing the activated state, and \bar{v} is the mean velocity of the activated complexes in the same coordinate, the frequency of crossing the barrier is given by \bar{v}/δ . If the number of activated complexes per unit volume lying in the length δ is c_1^\ddagger , then it follows that

$$\text{Rate of reaction} = c_1^\ddagger \bar{v} / \delta. \quad (103)$$

As indicated above, the activated complexes may be regarded as differing from normal molecules in the respect that one of the vibrational degrees of freedom is replaced by translational motion in the reaction coordinate. The complexes can, however, be treated as normal molecules with one very stiff vibrational mode, by expressing their concentration c_1^\ddagger at the top of the energy barrier in the form

$$c_1^\ddagger = c_1 (2\pi m_1 kT)^{1/2} \delta / h^3, \quad (104)$$

where m_1 is the effective mass of the activated complex in the reaction coordinate; k is the Boltzmann constant* and h the Planck constant. Comparison with (163), p. 873, shows that the factor $(2\pi m_1 kT)^{1/2} \delta / h^3$ is, in fact, the translational partition function of the activated complex in the reaction path; it may be taken as a measure of the probability of occurrence of the activated complex at the top of the barrier. The mean velocity \bar{v} of motion in one coordinate can be obtained by a method based on the Maxwell distribution law, similar to that used on p. 268 for the mean velocity in any direction. The value is found to be $(kT/2\pi m_1)^{1/2}$, and if this, together with (104), is introduced into (103), it is seen that the

* In this and subsequent equations k is the Boltzmann constant.

final result takes the remarkably simple form

$$\text{Rate of reaction} = c_{\ddagger}(kT/h). \quad (105)$$

It follows, therefore, that the effective rate of crossing the energy barrier by the activated complexes is equal to kT/h , which is a universal frequency, having the dimensions of time^{-1} ; its value is dependent only on the temperature and is independent of the nature of the reactants and the type of reaction.

If the specific reaction rate is represented by k_r^* , the reaction rate can be expressed in the familiar manner as

$$\text{Rate of reaction} = k_r c_A c_B \cdots, \quad (106)$$

and hence from (105) and (106),

$$k_r = \frac{kT}{h} \cdot \frac{c_{\ddagger}}{c_A c_B \cdots} = \frac{kT}{h} K^{\ddagger}, \quad (107)$$

where K^{\ddagger} is the constant for the equilibrium, viz., $A + B + \cdots \rightleftharpoons X^{\ddagger}$, which is supposed to exist between the reactants and the activated complex, assuming the system to behave ideally so that concentrations may be used in place of activities.

As seen on p. 882, for a standard state of 1 molecule per cc., the expression (191) for the equilibrium constant may be written in terms of the respective partition functions as

$$K = \frac{(Q^0)_{\text{products}}}{(Q^0)_{\text{reactants}}} e^{-\Delta E_0^0/RT}, \quad (108)$$

recalling that the translational contributions are for unit volume in every case. The equilibrium constant K^{\ddagger} may consequently be represented by

$$K^{\ddagger} = \frac{Q^{\ddagger}}{Q_A Q_B \cdots} e^{-E/RT}, \quad (109)$$

so that by (107)

$$k_r = \frac{kT}{h} \cdot \frac{Q^{\ddagger}}{Q_A Q_B \cdots} e^{-E/RT}. \quad (110)$$

The term ΔE_0^0 in (108) has here been replaced by E , the energy of activation, since the difference in energy between the activated complex, in its zero level, and the reactants, in their zero levels, i.e., ΔE_0^0 , is the energy of activation for the reaction when all the substances concerned are in their lowest energy levels. Since the activation energy does not vary greatly with temperature, it is evident that, as a first approximation, it is permissible to substitute ΔE_0^0 by E , as in (109) and (110). Attention should also be called to the fact that the partition functions refer to those in the standard state of 1 molecule per cc., and so k_r applies to concentrations in

* In order to distinguish it from the Boltzmann constant, the specific reaction rate will be represented in this section by the symbol k_r .

terms of this unit. Further, in evaluating Q^\ddagger for the activated complex, it should be remembered that one of the vibration frequencies must be treated as infinitely great, so that the corresponding contribution to the partition function is a factor of unity, and hence can be ignored.

An examination of (110) shows that it provides a basis for the calculation of the specific rate of a reaction merely from a knowledge of the physical properties of the reacting species. The values of Q_A and Q_B can be derived without difficulty, as shown in Chapter XI, and the activation energy E may be obtained from the potential energy surface as described above. Finally, it is possible to determine the dimensions and vibration frequencies of the activated complex from the potential energy surface, and so Q^\ddagger may be evaluated. As stated above, (110) solves the problem of the absolute calculation of reaction rates, in principle, but it is only in a very few simple cases that the potential energy surface can be obtained with sufficient accuracy to make the results reliable. Instead of applying the stringent test of (110), by calculating the actual specific rate, it can be tested both quantitatively and qualitatively in other ways, as will be seen below.

Comparison of Collision and Absolute Rate Theories.—Comparison of (110) with (94) shows that according to the absolute (or statistical) approach to the problem of reaction rates, the frequency factor A is given by

$$A = \frac{kT}{h} \cdot \frac{Q^\ddagger}{Q_A Q_B \cdots}, \quad (111)$$

and it is of interest to calculate the value of A by means of (111) for a simple reaction between two atoms, viz., A and B, the activated complex being the diatomic system AB^\ddagger . The partition functions of the atoms are made up of the translational contributions only, so that Q_A is $(2\pi m_A kT)^{3/2}/h^3$ and Q_B is $(2\pi m_B kT)^{3/2}/h^3$, the volume being unity in each case. A normal diatomic molecule has translational, rotational and one vibrational factor in its partition function, but for the activated complex the latter is unity, and so Q^\ddagger is the product of the translational contribution $[2\pi(m_A + m_B)kT]^{3/2}/h^3$, and the rotational contribution $8\pi^2 I kT/h^2$ (see p. 336). In the latter I is the moment of inertia of the activated complex, which in this case is equal to $\sigma^2 m_A m_B / (m_A + m_B)$, where σ is the mean diameter of the activated complex (p. 566). If these quantities are inserted in (111), the result is

$$A = \sigma^2 [8\pi R T (M_A + M_B) / M_A M_B]^{1/2}, \quad (112)$$

where R , equal to kN , is the gas constant per mole, and the M terms, equal to mN , are the molecular weights, N being the Avogadro number. It will be seen that this expression is identical with the collision number given on p. 1096 for a bimolecular reaction involving two different molecular species. It follows, therefore, that the statistical and simple collision theories lead to the same result, *provided the reactants are both atoms*.

The treatment given above for two atoms can be extended to reactions involving molecules, but as this becomes complicated a simplified procedure will be adopted. In the first place, the partition function for each type of energy will be assumed to consist of a number of equal terms, one for each de-

gree of freedom. If q_T , q_R and q_V are the translational, rotational and vibrational contributions, respectively, for one degree of freedom of each type, the complete partition function will then be represented by

$$Q = q_T^t q_R^r q_V^v, \quad (113)$$

where t , r and v are the numbers of the respective degrees of freedom. For the reaction between two atoms A and B, the values of Q_A and Q_B are each q_T^3 , while Q^\ddagger is $q_T^3 \times q_R^2$, the vibrational contribution being unity; hence (111) becomes

$$A = \frac{kT}{h} \cdot \frac{q_R^2}{q_T^3}. \quad (114)$$

As already seen, in these circumstances A becomes identical with the collision number Z , and so the quantity $(kT/h)q_R^2/q_T^3$ may be taken as equivalent to the latter. If now the foregoing, approximate, treatment is extended to the general case where A and B are nonlinear molecules containing a and b atoms, respectively, it is seen that

$$Q_A = q_T^3 q_R^3 q_V^{3a-6}; \quad Q_B = q_T^3 q_R^3 q_V^{3b-6}; \quad Q^\ddagger = q_T^3 q_R^3 q_V^{3(a+b)-7},$$

noting that the activated complex has one vibrational contribution less than a normal molecule. Substitution in (111) now gives

$$A = \frac{kT}{h} \cdot \frac{q_V^5}{q_T^3 q_R^3}, \quad (115)$$

and if this is represented by PZ , as in (102), and noting that Z is equivalent to $(kT/h)q_R^2/q_T^3$, it follows that the so-called probability factor P is

$$P = q_V^5/q_R^3. \quad (116)$$

At ordinary temperatures q_V is generally little different from unity, whereas q_R may be about 10 to 100, so that according to (116) the P factor in the reaction between two molecules, each containing three or more atoms, must be of the order of 10^{-5} to 10^{-10} . The more complex the reacting molecules the greater the value of q_R and so the smaller should be the P factor; this is in harmony with experimental findings. If the reacting molecules are diatomic, then it can be readily shown that $(q_V/q_R)^3$ corresponds to P , and since for such small molecules q_R is generally less than 10, it follows that the probability factor may well be of the order of 10^{-1} . This explains why the simple collision hypothesis accounts for the rates of the reactions $2\text{HI} = \text{H}_2 + \text{I}_2$ and $\text{H}_2 + \text{I}_2 = 2\text{HI}$, within a factor of about 10, and good agreement should also be obtained for reactions between an atom and a simple molecule. When the reactants are somewhat more complicated, as are those mentioned on p. 1098, the collision theory in its simple form may be expected to fail, as it in fact does.

Physical Significance of the Probability Factor.—The conclusions reached above may be put in another form which gives them some physical significance. When two polyatomic molecules react, the formation of the activated complex requires the disappearance of three translational and three rotational degrees of freedom of the reactants and the formation of five new vibrational modes in the activated state. It seems probable,

therefore, that the activated complex will be formed in a collision between the reactants only if the necessary rearrangement of the activation energy into the new degrees of freedom can occur. For atoms and simple molecules the energy transfer takes place readily, since few degrees of freedom are involved, and almost every collision with the requisite energy results in reaction. (In processes involving two atoms, however, a third body is usually required to carry off the excess energy, and so the reaction is not the simple one considered here.) With more complicated reactants the chances of forming the activated complex are small, since energy transitions in six degrees of freedom are necessary; only a small fraction of the activated collisions will lead to chemical reaction. The P factor is thus connected with the probability of the energy transformations that must occur if the activated state, which is an essential intermediate in all reactions, is to be formed, even when the total energy is available.

Attention should be drawn to the fact that if the collision theory were applied in the proper manner, it would give the same results as the statistical theory. According to the simple collision treatment, however, the reactants are regarded as rigid bodies with no internal degrees of freedom; this can only be justified if the reactants are atoms. If adequate allowance could be made in the calculation of the collision number for the necessary changes taking place in the distribution of internal energy, the results would presumably be the same as obtained from the absolute rate theory. In a sense, therefore, the latter treatment is an elaborated collision theory, for it provides a feasible approach to the problem of evaluating the number of encounters between molecules in which the conditions are such that reaction is possible.

Entropy of Activation.—The equations of the theory of absolute reaction rates may be put in another form which is useful in many respects. The equilibrium constant K^\ddagger may be expressed in terms of the free energy of activation in the standard state, ΔF^\ddagger , in the familiar form (p. 827) $-\Delta F^\ddagger = RT \ln K^\ddagger$, and since $\Delta F^\ddagger = \Delta H^\ddagger - T\Delta S^\ddagger$, it follows that

$$K^\ddagger = e^{\Delta S^\ddagger/R} e^{-\Delta H^\ddagger/RT}, \quad (117)$$

where ΔS^\ddagger and ΔH^\ddagger represent the difference in entropy and heat content, respectively, between the activated complex and the reactants, all referred to their standard states. If these are chosen correctly, it is possible to substitute for K^\ddagger , as given by (117), in (107), so that

$$k_r = (kT/h) e^{\Delta S^\ddagger/R} e^{-\Delta H^\ddagger/RT}, \quad (118)$$

which is almost identical with an equation deduced independently by V. K. La Mer (1933). The quantity ΔH^\ddagger may be assumed to be the same as the activation energy, and comparison of (118) with the expression $k_r = PZe^{-E/RT}$ shows that

$$PZ = (kT/h) e^{\Delta S^\ddagger/R}. \quad (119)$$

Since the theoretical collision number Z is virtually constant, and kT/h is constant, for different reactions, it is clear that the probability factor P is re-

lated to the entropy of activation ΔS^\ddagger . If this is positive, then P is large and the reaction will be "normal" or "fast," from the standpoint of the simple collision theory, but if ΔS^\ddagger is negative the reaction will be a "slow" one.

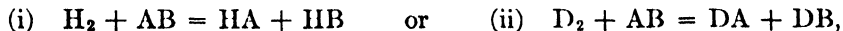
For a reaction involving two atoms or simple molecules there is a relatively small rearrangement of energy between the various degrees of freedom in the formation of the activated state, and in these circumstances ΔS^\ddagger has a small positive or negative value. The absolute rate and collision theories will then give results of the same order. On the other hand, when reaction occurs between complex molecules and there is a considerable rearrangement of energy among the degrees of freedom, there is a decrease of entropy in the formation of the activated complex, and the reaction rate will be lower than that calculated by the simple collision hypothesis. These qualitative conclusions have been confirmed by actual entropy calculations in certain cases where sufficient data are available for the entropy of the activated state to be determined from its partition function (p. 871). Thus, in the combination of hydrogen and iodine molecules, ΔS^\ddagger is almost zero, whereas in the bimolecular polymerization or ethylene, for which P is known to be about 10^{-3} , the value of ΔS^\ddagger is about -10 e.u. per mole, for the same standard state.

Free Energy of Activation.—One of the most significant conclusions to be drawn from the theory of absolute reaction rates is related to the free energy of activation. Since $-\Delta F^0$ is equal to $RT \ln K$, it is seen that K is equivalent to $e^{-\Delta F^0/RT}$, and if this substitution is made in (107), the result is

$$k_r = (kT/h)e^{-\Delta F^\ddagger/RT},$$

which is, of course, identical with (118). Since the factor kT/h is independent of the nature of the reaction, it follows that the specific rate of any reaction is determined by the factor $e^{-\Delta F^\ddagger/RT}$; hence it is the (standard) *free energy of activation which determines the rate of a reaction at a given temperature*. It was stated earlier that the energy (or heat) of activation is the essential factor, but this is an approximation, which is frequently useful.¹⁵

Activation Energy and Zero-Point Energy.—The difference in reactivity of the two isotopes of hydrogen may be explained by the difference in their zero-point energies; if hydrogen molecules react with a molecule AB, thus



it may be assumed that the activated complexes HHAB or DDAB will be formed. There is reason for believing that the potential energies of the two complexes are very similar, and hence the difference in the activation energies for the two reactions will depend on the difference in the energies of the two isotopic forms of molecular hydrogen; this is the difference in their zero-point energies (p. 157). If E^* is the energy of the activated complex, and ϵ_1 and ϵ_2 are the zero-point energies of molecular hydrogen and deuterium respectively, then the corresponding activation

energies E_1 and E_2 will be

$$E_1 = E^* - \epsilon_1 \quad \text{and} \quad E_2 = E^* - \epsilon_2. \quad (120)$$

The specific rate of a reaction may be expressed in the form $k = Ae^{-E/RT}$, and if it may be assumed, as is approximately true, that the frequency factor A is the same for the two isotopic reactions, it follows that

$$k_1/k_2 = e^{(\epsilon_1 - \epsilon_2)/RT}, \quad (121)$$

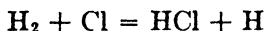
where $\epsilon_1 - \epsilon_2$ is the difference in the zero-point energies, which is about 1790 cal. The reaction between hydrogen and bromine is known to be complicated (p. 1079), but the rate-determining steps are probably



for the two isotopes. The ratio of the specific rates found experimentally is given by

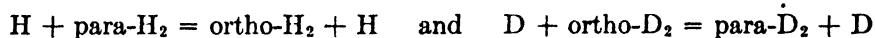
$$k_1/k_2 \approx e^{2130/RT},$$

which is in fair agreement with theoretical requirements. In the photochemical hydrogen-chlorine reaction, where the stage



probably determines the rate, the difference in activation energies for the reactions with the two isotopic forms of hydrogen is 1600 cal.

Study of the reactions



has brought out the important point that the activated complex, possessing to a great extent the properties of a real molecule, also has a zero-point energy. From the rates of the two reactions the difference in activation energy is found to be 510 cal., instead of the theoretical value of 1790 cal., and it is concluded that the zero-point energies of the complexes HHH and DDD differ by 1280 cal. (A. Farkas and L. Farkas, 1935). This difference is exceptionally high, but the complexes are here of an unusual isotopic nature.¹⁶

Theory of Termolecular Reactions.—The simple collision theory of reaction rates is even less successful for termolecular reactions than it is for bimolecular processes. As a rough approximation it may be stated that the number of ternary collisions in a gas is about 10^{-3} times the number of binary collisions at the same temperature and pressure (see p. 277); hence, allowing for unfavorable orientations, the relative rates of bimolecular and termolecular reactions at the same temperature should not be more than 1 to 10^{-4} . This would imply that a termolecular process should have an energy of activation about 5,500 cal. less at ordinary temperatures, or 18,000 cal. at 1000°K ., in order that it should have the same velocity as a bimolecular reaction. It would appear surprising, therefore, that so few termolecular reactions are known. The explanation

appears to lie in the fact that the formation of the activated complex, with its three degrees of translational and three of rotational freedom, from three individual molecules, with a total of nine degrees of translational and nine of rotational freedom, in the most general case, involves such a considerable rearrangement of energy that the probability of formation of the activated state is very small, even when the total requisite energy is available. This means that the P factor will be so minute that most termolecular processes will be so slow as to be undetectable. The same conclusion may be drawn from the fact that the formation of a single molecule, i.e., the activated complex, from three separate molecules is inevitably accompanied by a considerable restriction; hence the entropy of activation should have a large negative value, and the reaction should be correspondingly slow. It is significant that three-body collisions, involving two atoms and an energy removing molecule are not uncommon (p. 1064); in such cases the rearrangement of energy, or the decrease of entropy, will not be large, and provided the energy of activation is small, as it usually is for the combination of atoms, the rate may be expected to be appreciable.

One of the problems associated with termolecular reactions is to account for the observation that the specific rate of the reaction between nitric oxide and oxygen actually *decreases*, to some extent, with increasing temperature (M. Bodenstein, 1922). A possible explanation would be that the number of ternary collisions diminishes as the temperature is raised, but this appears to be improbable. An alternative view is based on a mechanism involving the intermediate formation of the complex $(\text{NO})_2$, which subsequently reacts with oxygen in a bimolecular process (cf. p. 1063). It will be seen from (41) that the observed specific reaction rate, k' , is equal to kK (p. 1063); the specific rate k will presumably increase with increasing temperature, but if the association reaction $2\text{NO} \rightarrow (\text{NO})_2$ is exothermic, as it probably is, the equilibrium constant K will decrease. If the activation energy for the reaction between $(\text{NO})_2$ and oxygen is less than the heat change accompanying the formation of $(\text{NO})_2$, k will increase less rapidly than K decreases, and consequently k' will be diminished as the temperature is raised (M. Bodenstein, 1935).

An alternative interpretation of the results of the oxygen-nitric oxide reaction, which does not necessarily invalidate the foregoing mechanism, is provided by application of the absolute reaction rate theory, in the form of (110). By inserting the expressions for the various partition functions of reactants and activated complex, at moderate temperatures, the equation reduces to the form

$$k = BT^{-3}e^{-E/RT},$$

where B is approximately independent of temperature.* It can be seen

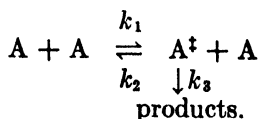
* According to this expression the frequency factor A is equal to BT^{-3} , and hence is markedly temperature dependent. Actually A is temperature dependent, to some extent, for reactions of all types, irrespective of their order, but the change in $e^{-E/RT}$ with temperature is usually so large that the variation of A can be neglected.

that if E is small, as it is for the reaction under consideration, the increase in $e^{-E/RT}$ may well be less than the decrease in T^{-3} , as the temperature is raised, and so the specific rate k may actually diminish (H. Gershinowitz and H. Eyring, 1935).

Unimolecular Reactions.—Discussion of unimolecular reactions has been left to the last as they involve special factors. For unimolecular changes, which are also of the first order, the time taken to reduce the concentration, or pressure, of reactant by a definite fraction is independent of that concentration, or pressure. For a given mass of gas, the amount decomposed in unit time, i.e., the rate of reaction, will, therefore, be independent of the volume, and hence of the pressure. It is evident that chemical reaction cannot be a *direct* consequence of collision, for the number of collisions in a given mass of gas will depend on the volume occupied. Measurements of the velocities of first order homogeneous gas reactions have shown that quite considerable energies of activation are required before the molecules can decompose, values of 40,000 to 60,000 cal. per mole being common. It was suggested by J. Perrin (1919) that the requisite energy is obtained by the absorption of infrared radiation, but this view, like the general radiation theory of reaction velocities, meets with many difficulties (cf. p. 1090). An alternative theory, developed by J. A. Christiansen and H. A. Kramers (1923), was to the effect that unimolecular reactions take place by a chain mechanism; it was considered that the product of the reaction was able to transfer its energy to a molecule of reactant and so activate it. There are many objections to this hypothesis and it need not be considered further here. The view now widely accepted concerning the mechanism of unimolecular reactions is based on a suggestion of F. A. Lindemann (1922): the reacting molecules are supposed to acquire energy of activation as a result of collision with others, but they decompose only when a certain period of time has elapsed after impact. If the average time interval between activation by collision and decomposition is large compared with that between two collisions, the majority of the activated molecules will lose some of their energy by another collision before they have time to decompose, and so only a small fraction of the active molecules present at any instant will succeed in reacting. The Maxwell distribution law shows that the number of activated molecules must be a definite fraction of the whole number irrespective of the volume occupied, and if the activated molecules decompose at a definite rate, the number decomposing in a given time will be a definite fraction of the total number, also independent of the volume, or pressure. This is, in fact, the condition for a first order reaction, as explained above.

The same conclusion may be reached in a somewhat different manner. If the normal molecule of reactant is represented by A and the activated molecule by A^\ddagger , then the stages of the unimolecular reaction may be

written in the form



The rate of formation of activated molecules is proportional to the square of the concentration of normal molecules, which is practically the same as the total concentration c ; that is,

$$\text{Rate of activation} = k_1 c^2. \quad (122)$$

The rate of deactivation of active molecules by loss of energy to normal molecules in collisions is proportional to the concentration of active molecules a and normal molecules c ; hence,

$$\text{Rate of deactivation} = k_2 a c. \quad (123)$$

The active molecules decompose, when they are not deactivated, at a rate represented by $k_3 a$, and so for a stationary concentration of active molecules,

$$k_1 c^2 = k_2 a c + k_3 a; \quad (124)$$

$$\therefore \text{Rate of reaction} = k_3 a = k_1 c^2 / (k_2 c / k_3 + 1). \quad (125)$$

If c is large then $k_2 c / k_3$ is large in comparison with unity, and the latter may be neglected, giving

$$\text{Rate of reaction} = (k_1 k_3 / k_2) c. \quad (126)$$

The rate is thus directly proportional to the concentration of the reactant and hence the reaction is first order. The Lindemann mechanism, therefore, although it involves the concept of activation already outlined, can account for the kinetics of unimolecular reactions.

At low pressures the time between two successive collisions is increased, and eventually a pressure will be reached when the interval is equal to the time elapsing between activation of a molecule by collision and its decomposition. At this pressure, and at all lower values, every molecule acquiring energy of activation will decompose before it can become deactivated. The rate of reaction will thus be equal to the rate of formation of activated molecules, and since this depends on the square of the molecular concentration [equation (122)], the unimolecular process should become kinetically of the second order. This result may be derived by assuming c to be small in (125), as it would be at low pressure; then $k_2 c / k_3$ is negligible in comparison with unity and it follows that

$$\text{Rate of reaction} = k_1 c^2, \quad (127)$$

which is the condition for a second order process.

The experimental verification of this expectation is one of the strongest arguments for the mechanism proposed. In almost every homogeneous

unimolecular gas reaction studied it has been observed that below a certain pressure the process ceases to be of the first order, and the specific rate decreases as the pressure is lowered still further. In Table 183 are

TABLE 183. DECOMPOSITION OF AZOMETHANE AT 330° C.

<i>p</i> cm.	39.3	23.5	14.4	5.65	3.33	1.44	0.75
<i>k</i> × 10 ³	2.82	2.82	2.65	2.13	1.76	1.31	1.10

given the specific rates (*k*) for the decomposition of azomethane for various initial pressures (*p*) at 330° c. (H. C. Ramsperger, 1927). In some instances, notably the decomposition of nitrogen pentoxide, it is necessary for very low pressures, viz., 0.05 mm. of mercury at 35° c., to be attained before the falling off in reaction velocity is detected; at a pressure of 0.001 mm. the apparent order of the decomposition of nitrogen pentoxide is 1.8. It is of interest to note that although the reactions become of the second order kinetically at low pressures, they are still unimolecular in the sense that each *chemical* act involves only one, and not two, molecules. This is a further illustration of the distinction already made (p. 1045) between the order of the reaction and its molecularity.

Calculation of Unimolecular Reaction Rates.—Application of the simple collision theory to the concept of activation outlined above suggests that the maximum specific rate of a unimolecular reaction would be given by $Ze^{-E/RT}$, where the collision number *Z* is now defined by $4\sigma^2(\pi RT/M)^{1/2}$, as given on p. 1096. It is found, however, that the observed specific rates for most unimolecular reactions are several powers of ten greater than the values calculated in this manner. An attempt was made to overcome this difficulty by postulating that energy from several internal degrees of freedom of the molecule, and not merely in two square terms (p. 1090), can contribute to the energy of activation. Some confirmation of this view appeared to be provided by the fact that the number of such degrees of freedom which had to be invoked increased with increasing complexity of the reacting molecule. However, there is the fundamental difficulty of explaining why it is not necessary to make a similar postulate in connection with bimolecular processes, even when relatively complex molecules are involved.

Once again, a more satisfactory interpretation of the rates of unimolecular reactions is provided by the theory of absolute reaction rates. If a molecule A undergoes unimolecular decomposition, it must first acquire the necessary energy in the proper degrees of freedom, in a suitable collision, and thus form the activated state A[‡]. The latter will not differ greatly from the reacting molecule, and the only energy transition involved is from one vibrational degree of freedom in A to translation along the reaction coordinate in A[‡]. The probability of the formation of the activated state may thus be high, and hence the specific rate may be greater than calculated from the simple collision theory. The same conclusion is reached by a consideration of the entropy of activation.

In general, the activated complex A^\ddagger in a unimolecular process will have a looser structure than the reactant A , and hence there will be a greater randomness of motion in the activated state. The formation of the latter will thus be accompanied by an increase of entropy, i.e., ΔS^\ddagger has an appreciable positive value; the specific rate will then be larger than $Ze^{-E/RT}$, as actually found. The (standard) entropies of activation for a number of unimolecular reactions* calculated from the observed specific rates by (118), are recorded in Table 184; they are seen to be positive, as expected.

TABLE 184 ENTROPIES OF ACTIVATION IN UNIMOLECULAR GAS REACTIONS

Decomposition of	A (sec. ⁻¹)	E	ΔS^\ddagger
CH_3OCH_3	1.6×10^{13}	58.5 kcal.	2.5 e.u.
CH_3N_3	3.0×10^{13}	43.5	8.2
$\text{CH}_3\text{N}=\text{NCH}_3$	8.0×10^{13}	50.0	10.8
$(\text{CH}_3)_2\text{O}$	6.3×10^{14}	52.0	5.5
$\text{Si}(\text{CH}_3)_4$	1.7×10^{14}	78.8	2.2

Another aspect of the subject is provided by a consideration of (110), using the approximate procedure (p. 1102). It can be readily shown, since the activated state differs from the reacting molecule by one vibrational mode, that the frequency factor A for a unimolecular reaction is equivalent to $(kT/h)q_v$. If the value is to be the same as given by the simple collision theory, it should be approximately $(kT/h)q_R^2/q_1^3$, and so that required by the statistical theory is greater by a factor of $q_1^3/q_R^2q_v$. This quantity is very large, and so the frequency factor A calculated from the absolute reaction rate theory is much greater than given by the simple collision theory.

It should be pointed out that not all unimolecular reactions have high velocities; the frequency factors in Table 184 are of the order of 10^{14} sec.⁻¹, but in some cases values of 10^1 or less have been observed. This is sometimes due to the fact that the activated complex has a more rigid structure than the initial state, and hence ΔS^\ddagger is negative. Another factor which accounts for unimolecular reactions having lower velocities than might be expected from (110) is that not all the molecules passing through the activated state proceed to decomposition. This possibility is sometimes stated in the form that the "transmission coefficient" is less than unity. The frequency of passage of activated complexes over the barrier is thus less than kT/h . There are several reasons for the small transmission coefficient; among these mention may be made of the possibility that the concentration of activated complexes is less than the equilibrium value because of the slowness of the transfer of energy among various degrees of freedom.

The views described above provide an outline for the treatment of unimolecular reactions, although various slight modifications, e.g., concerning the activation of special degrees of freedom, have been made.

* The standard entropy of activation, like the specific rate, of a unimolecular reaction is independent of the standard state, i.e., of the concentration units.

The most serious matter to consider, however, is the possibility that chain mechanisms may be concerned. It was seen on p. 1085 that a chain reaction involving free radicals might be kinetically of the first order, and free radicals may well be formed in the decomposition of some of the complex molecules taking part in apparently unimolecular reactions. It has been established that in certain cases, e.g., the first order decomposition of methylamine and propylamine, chains play a part in the reaction, but it is not certain as to whether they have any marked influence on the kinetics. It is probable, however, that first order gas reactions, which might be imagined to be the most simple of all chemical changes, are actually complicated processes and may well involve several different mechanisms.¹⁷

REACTIONS IN SOLUTION

Reactions in Gas Phase and in Solution.—A number of reactions are known for which the velocities are similar in the gas phase and in solution; in such instances there can be little doubt that the same mechanisms are effective under the two different conditions. Examples are the decomposition of chlorine monoxide, the decomposition of ozone catalyzed by chlorine, and the thermal decomposition of ethylene iodide, all in carbon tetrachloride solution. Reference has been made (p. 1050) to the fact that nitrogen pentoxide decomposes at the same rate, within a factor of two, in a number of solvents as it does in the gaseous state, and the same applies to the isomerization of *d*-pinene. Several so-called "slow reactions," e.g., between triethylamine and ethyl iodide, and between ethyl alcohol and acetic anhydride, also have specific rates of the same order in solution as in the gas phase.

In deriving (107) in the form $k_r = (kT/h)K^\ddagger$ from the theory of absolute reaction rates, it was assumed that the substances concerned behaved ideally, so that the concentrations c could be substituted for activities. Strictly speaking c is equal to a/α , where a is the activity and α the activity coefficient,* and hence (107) may be written as

$$k_r = \frac{kT}{h} \cdot \frac{a_\ddagger}{a_A a_B \cdots} \cdot \frac{\alpha_A \alpha_B \cdots}{\alpha_\ddagger} = \frac{kT}{h} K_0^\ddagger \frac{\alpha_A \alpha_B \cdots}{\alpha_\ddagger}, \quad (128)$$

where K_0^\ddagger is the true equilibrium constant. The quantity $(kT/h)K_0^\ddagger$ represents the specific rate of the reaction in an ideal system, e.g., gases at low pressures or a dilute solution, when the activity coefficients are unity; it may therefore be replaced by k_0 , so that (128) becomes

$$k_r = k_0 \frac{\alpha_A \alpha_B \cdots}{\alpha_\ddagger}. \quad (129)$$

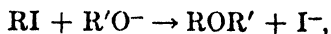
* This unusual symbol α is used for the activity coefficient for it will be referred to an unusual standard state (p. 1112).

For most reactions occurring in the gas phase at moderate pressures, the activity coefficient factor is close to unity, for the departure from ideal behavior is not large; it is then possible to represent the specific rate k_g of the gaseous reaction by k_0 . If k_s is the rate of the same reaction measured in solution, then by (129)

$$k_s = k_g \frac{\alpha_A \alpha_B \cdots}{\alpha_{\ddagger}}, \quad (130)$$

where the activity coefficients have the unusual significance of applying to the reactants A, B, etc., and the activated complex in *solution*, with reference to the *ideal gas* at unit concentration as the standard state. There is no simple method for evaluating these activity coefficients, but an approximate indication can be obtained from solubility measurements made on the gases. From such data it appears that α is of the order of 10^2 for a number of substances. It follows, therefore, from (130), that for a bimolecular reaction k_s should be approximately 100 k_g , while for a unimolecular reaction k_s and k_g should be roughly equal. The latter result may be derived in an alternative manner from (130); for a unimolecular reaction this reduces to the simple form $k_s = k_g(\alpha_A/\alpha_{\ddagger})$. Since the reacting substance and the activated state are not very different (cf. p. 1109), the ratio $\alpha_A/\alpha_{\ddagger}$ must be of the order of unity; hence k_s and k_g will be approximately equal. It is seen, therefore, from the foregoing discussion, that the specific rate of a unimolecular or bimolecular reaction will not be very different in the gas phase and in solution. This is in agreement with the results recorded above. However, should there be any interaction between a reactant and the liquid solvent, the activity coefficients may be so changed as to cause the rate of the process in solution to differ considerably from that in the gas phase.

The replacement of the equilibrium constant by the ratio of partition functions, as in (110), is not permissible for equilibria in solution, but the fact that the rate of a given reaction is of the same order in solution as in the gas phase makes the same general conclusions applicable to both cases. For example, it is to be expected that for reactions in solution involving simple molecules the rates will be given approximately by $k = Ze^{-E/RT}$, where Z is the collision number calculated as for gases. It appears that, in general, all reactions involving a molecule and an ion, e.g., between an alkyl iodide and an alkoxy or phenoxy ion, viz.,



give calculated specific rates within an order of ten of the observed values (G. H. Grant and C. N. Hinshelwood, 1933; E. A. Moelwyn-Hughes, 1933). Similar satisfactory agreement has been obtained in connection with catalyzed reactions, such as the mutarotation of beryllium benzoyl camphor and of glucose, which are kinetically of the first order but are actually bimolecular processes involving the reactant and a

catalyst, usually a simple ion. On the other hand, for reactions between two relatively large molecules, e.g., between substituted benzoyl chlorides and anilines, between triethylamine and ethyl iodide, and between ethyl alcohol and acetic anhydride, the observed specific rate is much less than required by the simple collision equation.¹⁸

Influence of Solvent.—The effect of solvent on reaction velocity was discovered by M. Berthelot and P. de St. Gilles (1862), and although much work has been done on the subject it still remains a highly perplexing problem in physical chemistry. The first systematic study was made by N. A. Menshutkin (1887 *et seq.*) who examined the reaction between triethylamine and ethyl iodide in a large number of solvents; the velocity constants ranged from 0.00018 in hexane to 0.133 in benzyl alcohol at 100° c. It was thought at one time that there was a parallelism between the dielectric constant of the solvent and the reaction velocity in it, but the relationship is only very approximate and numerous exceptions are known. Further, not all reactions are influenced in the same manner by a series of solvents. Reference has been made to the decomposition of nitrogen pentoxide which occurs at almost the same rate in a number of media, and the same appears to be true for the decomposition of hexaphenylethane. For the reaction between ethyl alcohol and acetic anhydride, on the other hand, the solvent influence is almost the reverse of that observed for the reaction of triethylamine and ethyl iodide (F. G. Soper, 1931). Some of the results obtained for reactions of the three types in various solvents are given in Table 185. In spite of the com-

TABLE 185. INFLUENCE OF SOLVENT ON REACTION VELOCITY

Solvent	Acetic Anhydride and Ethyl Alcohol at 50°	Decomposition of Nitrogen Pentoxide at 20°	Triethylamine and Ethyl Iodide at 100°
Hexane	0.0119	—	0.00018
Carbon tetrachloride	0.0113	0.000235	—
Chlorobenzene	0.00533	—	0.023
Benzene	0.00462	—	0.0058
Chloroform	0.00404	0.000274	—
Anisole	0.00293	—	0.040
Nitrobenzene	0.00245	—	70.1

plexity of the solvent influence, the following general rules (F. G. Soper, 1929), appear to provide a useful guide. If the reaction is one in which the products have a higher internal pressure (p. 479) than the reactants, then it is accelerated by solvents of high internal pressure; if the products are of lower internal pressure, however, the reaction will be retarded by these solvents. Since internal pressure and polarity often run parallel, a polar solvent might be expected to accelerate a process in which a polar substance is the product and retard one in which a nonpolar substance is formed. In the reaction between triethylamine and ethyl iodide, and in

other analogous processes, the product is a salt and hence more polar than the reactants; the rate should, therefore, be increased as the internal pressure or polarity of the solvent increases, as is, in fact, generally the case. On the other hand in the acetylation of alcohol, or in the decomposition of a sulfonium salt into an alkyl halide and sulfide, the products are less polar than the reactants, and the solvent effect should be reversed; this is also in approximate agreement with observation (Table 185). In the decomposition of nitrogen pentoxide the difference in internal pressure between the reactant and the products of the rate-determining step, possibly nitrogen trioxide and oxygen, is presumably small, so that the rate is almost independent of the nature of the solvent.

If either the reactants or the activated complex interacts with the solvent, there may be a considerable effect on the reaction rate; the exceptionally high value for the triethylamine-ethyl iodide reaction in nitrobenzene (Table 185) is probably to be accounted for in this manner. The effect of interaction with the solvent is to lower the potential energy of the substance by an amount equal to the energy of solvation. If one or both the reactants is solvated, while the activated complex is not, the decrease in potential energy of the former means that the necessary activation energy is increased; the reaction rate is consequently diminished. If the activated complex is solvated, but the reactants are not, the lowering of the potential energy means an equivalent decrease in the energy of activation of the reaction; the specific rate is then greater than normal. In the event that a reactant and the activated complex are both solvated, the over-all effect on the activation energy and the specific rate may be small.

Reactions Involving Ions.—Although the activity coefficients α in (130) cannot be determined readily, since they are for constituents of a solution referred to ideal gas as the standard state, it is nevertheless possible to modify the result so as to be applicable to a single medium. If f is the usual activity coefficient for a solute, with reference to the ideal molar solution in the same solvent as the standard state, then α is equal to βf , where β is a constant for a given substance in a particular solvent. Making this substitution, for a bimolecular reaction, (130) becomes

$$k_s = k_o \frac{\beta_A \beta_B}{\beta_{\ddagger}} \cdot \frac{f_A f_B}{f_{\ddagger}}. \quad (131)$$

Since the β values are all constants for the reactants in the chosen solvent, the factor involving the β 's may be combined with k_o to give a constant k_o^s , applicable to the given solvent; hence, (131) takes the form

$$k_s = k_o^s \frac{f_A f_B}{f_{\ddagger}}, \quad (132)$$

or taking logarithms, and omitting the symbol s , it is found that

$$\log k = \log k_o + \log f_A f_B / f_{\ddagger}. \quad (133)$$

This expression is known as the Brønsted-Bjerrum equation; it was derived in different ways by J. N. Brønsted (1922) and N. Bjerrum (1924), and others. The term $f_A f_B / f_{\ddagger}$, involving the activity coefficients, is called the **kinetic activity factor**, for its significance is identical with that of the kinetic factor \mathfrak{F} considered on p. 826. At infinite dilution this factor becomes equal to unity, and hence k_0 in (133) is the "observed" specific rate in such a solution; the difference between this and the value obtained in a solution of appreciable concentration depends on the variation of the activity factor with concentration.

The simple derivation of (133) given by Bjerrum is based on similar principles to those used in the absolute rate theory, but involves less detail. It is supposed that the reactants A and B are in equilibrium with the activated complex X^\ddagger , so that

$$K = a_{\ddagger} / a_A a_B, \quad (134)$$

while the reaction rate is proportional to the *concentration* of the activated complex, i.e., $k_{\ddagger} c_{\ddagger}$. Replacing the activities by the products of the concentrations and activity coefficients, it is found from (134) that

$$c_{\ddagger} = K c_A c_B \frac{f_A f_B}{f_{\ddagger}},$$

and hence,

$$\text{Rate of reaction} = k_{\ddagger} c_{\ddagger} = k_{\ddagger} K c_A c_B \frac{f_A f_B}{f_{\ddagger}}. \quad (135)$$

The rate of reaction can also be expressed in the usual manner by $k c_A c_B$, where k is the observed specific rate, and combination with (135) gives

$$k = k_{\ddagger} K f_A f_B / f_{\ddagger} = k_0 f_A f_B / f_{\ddagger}, \quad (136)$$

which is identical with (132) or (133).

If the reactants and products are all nonelectrolytes, either in the gaseous state or in solution, the kinetic activity factor will be almost independent of concentration, and so the observed velocity "constant" k will also remain virtually constant. The reaction velocity will thus be very closely proportional to the concentration of the reactants over a considerable range. If ions are involved, however, the value of k may be constant in a particular solution, but it will vary with the total ionic concentration, whether due to the reactants themselves or to added electrolytes.

Primary Salt Effect.—The effect of electrolytes can be divided into two categories, known as the **primary** and **secondary salt effects**; the former refers to the influence of the electrolyte concentration on the activity coefficients f_A , f_B and f_{\ddagger} , whereas the latter is concerned with actual changes in the concentration of the reacting ions resulting from the addition of electrolytes. Both effects are of importance in the study of ionic catalysis in solution (p. 1138), but the primary salt effect is involved in

noncatalytic reactions and this aspect of the subject will be considered here.

The variation of activity coefficient with concentration in dilute solutions may be represented by the simple limiting form of the Debye-Hückel equation (p. 959); thus for an ion of charge z , in aqueous solution at 25° c.,

$$-\log f = 0.51z^2 \sqrt{\mu}, \quad (137)$$

where μ is the ionic strength of the solution (p. 956). It follows, therefore, that

$$\log \frac{f_A f_B}{f_{\ddagger}} = -0.51 \sqrt{\mu} (z_A^2 + z_B^2 - z_{\ddagger}^2), \quad (138)$$

where z_A , z_B and z_{\ddagger} are the charges carried by A, B and X_{\ddagger}^{\ddagger} , respectively. Since the activated complex is made up of A and B, its charge will be the algebraic sum of those of A and B; that is $z_{\ddagger} = z_A + z_B$, the appropriate sign being included, and so (138) becomes

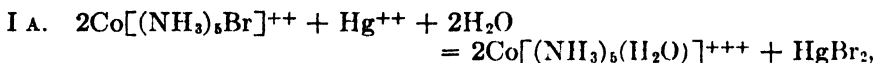
$$\log \frac{f_A f_B}{f_{\ddagger}} = 0.51 \sqrt{\mu} (2z_A z_B) = 1.02 z_A z_B \sqrt{\mu} \quad (139)$$

at 25° c. Substitution in (134) then gives

$$\log k = \log k_0 + 1.02 z_A z_B \sqrt{\mu}. \quad (140)$$

The plot of the observed values of $\log k$ against $\sqrt{\mu}$ for aqueous solutions at 25° c. should thus be a straight line of slope $1.02 z_A z_B$.

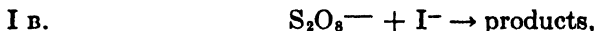
Measurements of the velocities of various reactions involving ions have provided clear confirmation of the Brønsted relationship in the form of (140). The variation of the specific rate k with ionic strength should depend on the magnitude and sign of $z_A z_B$, and three main types of behavior may be distinguished. In the first $z_A z_B$ is positive, and so k should increase with increasing concentration of ions; examples of this nature are the following:



in which the rate-determining stage is the bimolecular process



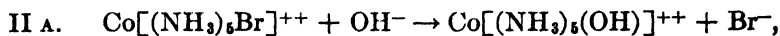
so that $z_A z_B$ is 4.



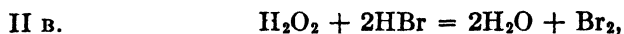
which determines the rate of reaction between persulfate and iodide ions; here $z_A z_B$ is + 2, although both ions have a negative valence.

In the second type of reaction $z_A z_B$ is negative, so that the velocity

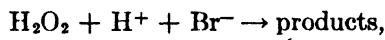
constant decreases with increasing ionic strength; examples are as follows:



giving $z_A z_B = -2$.



where the stage



for which $z_A z_B$ is -1 , controls the rate of the reaction.

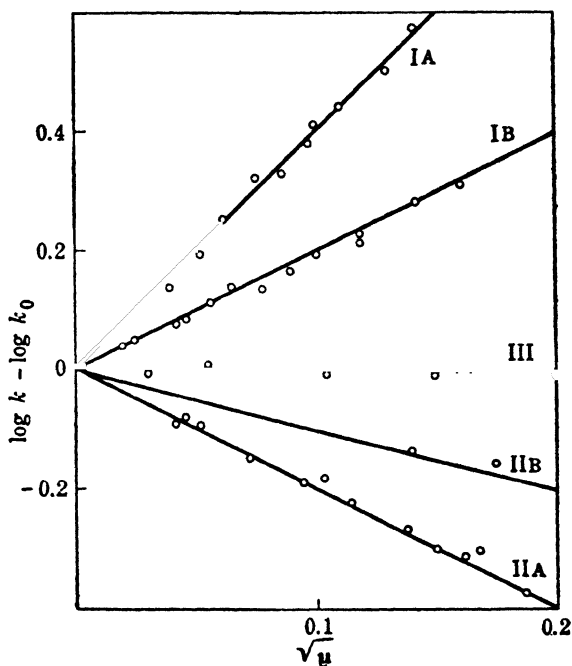
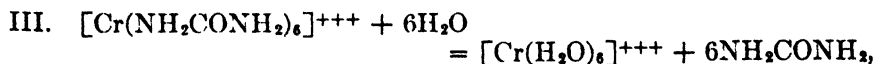


Fig. 250. Dependence of reaction rates on ionic strength

In the third category are those for which $z_A z_B$ is zero, so that in dilute solution, at least, the ionic concentration has little or no effect on the velocity constant. As an illustration the reaction below may be quoted.

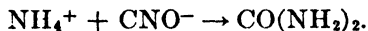


in which the hexa-urea-chromium ion is decomposed by water; the essential stage here appears to be the addition of a neutral water molecule to a

tervalent ion so that $z_A z_B$ is zero. A number of catalytic reactions of this type have been studied and reference to them will be made later (p. 1138).

In Fig. 250 the values for $\log k - \log k_0$ for solutions of different electrolyte concentrations have been plotted against the corresponding values of \sqrt{y} ; the full lines show the results to be expected from (140), while the points give the corresponding experimental values. The agreement is seen to be very striking. It must be remembered that since (140) involves the Debye-Hückel relationship (137), which is strictly applicable at high dilutions, it can be expected to hold only for dilute solutions; in relatively concentrated solutions, especially if ions of high valence are present, deviations are observed.

The effect of electrolytes on the isomeric change of ammonium cyanate to urea has been employed, in conjunction with the Brønsted equation, to verify the suggestion that the process is bimolecular (p. 1059) and involves ammonium and cyanate ions, viz.,



If the rate-determining stage is as depicted, then $z_A z_B$ is -1 , and the observed velocity should decrease with increasing ionic concentration; this has been found to be the case, the slope of the straight line plot of $\log k$ against \sqrt{y} being in agreement with theory.²⁰

KINETICS OF HETEROGENEOUS REACTIONS

Gas Reactions on Solid Surfaces.—The study of the kinetics of reactions between gases taking place on the surfaces of solids has led to the discovery that in addition to reactions of apparently simple orders, there are others of zero or fractional orders. In some cases the reaction is inhibited, instead of accelerated, by one of the reactants, whereas in others the products may have a retarding influence. The first satisfactory interpretation of these results was given by I. Langmuir (1916), based on the suggestions made in connection with his theory of adsorption (p. 1198). The treatment of Langmuir, as given below, employs simple kinetic considerations; the same conclusions can, however, be reached by application of the theory of absolute reaction rates.

Single Reacting Gas.—In a system consisting of a solid surface and a gas, the molecules of the latter will strike the surface and, as a general rule, will "condense," that is, adhere for an appreciable period. As a result of thermal agitation, however, the gas molecules will "evaporate," that is, leave the surface, from time to time, and eventually an equilibrium is reached when the rates of evaporation and condensation are equal. The mass of a given gas, and hence the number of molecules (μ),* striking 1 sq. cm. of surface per sec. is proportional to the pressure (p) of the gas.

* The quantity μ derived on p. 278 refers to the mass of gas, but the same symbol is used here for the number of molecules striking the surface.

If α is the fraction of the molecules striking the surface which adhere, then $\alpha\mu$ molecules condense on each sq. cm. of available surface per sec. If θ is the fraction of the total surface covered with gas molecules at any instant, then $1 - \theta$ is the fraction of the surface which is bare; assuming that only a single layer of gas molecules can form on the solid (cf. p. 1198), the actual rate of condensation will be $(1 - \theta)\alpha\mu$ molecules per sq. cm. of total surface per sec. The rate of evaporation will be proportional to the number of molecules on the surface, that is, to the area covered, θ , and so it may be represented by $\nu\theta$, where ν is a constant for the given gas and surface. At equilibrium, the rates of condensation and evaporation of the gas molecules will be equal, so that

$$(1 - \theta)\alpha\mu = \nu\theta; \quad (141)$$

$$\therefore \theta = \frac{\alpha\mu}{\alpha\mu + \nu}. \quad (142)$$

It is now necessary to consider the question of the decomposition of the gas on the surface. If the assumption is made that the reaction velocity is small in comparison with the rates at which molecules condense on and evaporate from the surface, the fact that some of them decompose will not affect appreciably the equilibrium attained between the molecules on the surface and those in the gas phase, as expressed by (141).

In general, chemical reaction may be considered to occur between m adjacent molecules, followed by immediate evaporation of the products, so that the rate of reaction is given by

$$\frac{dx}{dt} = \nu_1\theta^m, \quad (143)$$

where ν_1 determines the rate of evaporation of the products, θ being a measure of the surface concentration of the reacting molecules. In the simplest case m is equal to unity, and so

$$\frac{dx}{dt} = \nu_1\theta. \quad (144)$$

Substituting the value of θ from (142) in (144), and utilizing the fact that μ is proportional to the pressure p , then

$$\frac{dx}{dt} = \frac{k_1 p}{k_2 p + 1}, \quad (145)$$

where k_1 and k_2 are constants, proportional to $\nu_1\alpha/\nu$ and to α/ν respectively. According to the assumption made above, that only one molecule of reacting gas is involved in the reaction, the process is unimolecular, but examination of (145) shows that it is not a simple first order reaction.

If the surface is sparsely covered, as it would be at relatively high temperatures, when the molecules do not condense easily, i.e., α is small,

but evaporate readily, i.e., ν is large, θ will be small in comparison with unity; (141) will then become $\alpha\mu = \nu\theta$, and combination with (144) gives

$$\frac{dx}{dt} = kp. \quad (146)$$

A unimolecular heterogeneous reaction is thus kinetically of the first order only when the surface of the solid is sparsely covered with molecules. Examples of such reactions are the decomposition of phosphine on glass, porcelain, silica and tungsten surfaces; of arsine on glass; of nitrous oxide on gold; of hydrogen iodide on platinum; and of formic acid vapor on glass, platinum, titanium oxide and other surfaces.

When the surface is almost completely covered, that is to say, when θ approaches unity, another approximation is possible. By setting θ equal to unity on the right-hand side of (144), it is seen that

$$\frac{dx}{dt} = \text{constant}, \quad (147)$$

since ν_1 is a constant. The rate of reaction will thus be independent of the pressure, and the process will be kinetically of zero order. It has actually been found in the decomposition of hydrogen iodide on a gold surface that the rate of decomposition remains almost unchanged in spite of a four-fold change of pressure; somewhat similar results have been observed when ammonia is decomposed on the surface of tungsten. In these instances it is probable that the reaction is unimolecular and occurs on a surface almost completely covered with the reacting molecules. It is easy to see why the reaction should be of zero order. If the surface is virtually covered, change of pressure will have relatively little effect on the surface concentration of reacting molecules and hence the rate of decomposition will remain almost constant. The decomposition of phosphine on a molybdenum surface is kinetically of the first order at low pressures but becomes of zero order as the pressure is increased; at the higher pressures the adsorption of the phosphine is increased and the surface is probably completely covered.

It can be seen that the two extreme cases of (145) are given by (146) as $dx/dt = kp$, and by (147) as $dx/dt = kp^0$, so that the general form of (145)* would be

$$\frac{dx}{dt} = kp^n, \quad (148)$$

where n lies between 0 and 1. The decomposition of stibine actually satisfies this type of equation with n equal to 0.6, although the reaction is probably unimolecular. As the temperature is raised the observed order tends toward unity in agreement with expectation.

* Equations (146) and (147) may be obtained from (145) by neglecting k_p in comparison with unity and *vice versa*, respectively.

Retardation by Reaction Products.—In some reactions the products are capable of retarding the reaction; this effect appears to result from the products being strongly held, that is, preferentially adsorbed (see p. 1194), on the surface. A simple case will be taken to determine the kinetics to be expected. Suppose the reacting gas occupies only a small fraction of the surface; most of the fraction θ of the solid covered by gas molecules is occupied by the product. When equilibrium is reached between molecules of *product*, which condense on and evaporate from the surface, the condition is

$$(1 - \theta)\alpha_1\mu_1 = \nu_1\theta; \quad (149)$$

$$\therefore 1 - \theta = \frac{\nu_1}{\alpha_1\mu_1 + \nu_1}, \quad (150)$$

where μ_1 is proportional to the pressure of the *product* in the gas phase. It is convenient now to suppose that every molecule of reactant striking the *bare* parts of the surface decomposes, the reaction being unimolecular; then

$$\frac{dx}{dt} = k\mu(1 - \theta), \quad (151)$$

where μ refers to the reactant. Substituting the value for $1 - \theta$ given by (150) in (151), it follows that

$$\frac{dx}{dt} = \frac{k\mu\nu_1}{\alpha_1\mu_1 + \nu_1} = \frac{k_1p}{k_2p_1 + 1}, \quad (152)$$

p being the pressure of the reactant and p_1 that of the product; as the latter accumulates, therefore, it retards the reaction.

The rate of decomposition of nitrous oxide on a platinum surface at about 1000° c. is retarded by oxygen, one of the products, and the velocity is given by an equation of the form of (152) with p equal to the pressure of nitrous oxide, and p_1 that of oxygen. It appears, therefore, that most of the platinum surface is covered with oxygen which, by preventing access of the nitrous oxide, retards the decomposition process.

If the reaction product is very strongly absorbed on the surface, its rate of condensation, determined by α_1 , will be great in comparison with the rate of evaporation, dependent on ν_1 . Under these conditions α_1/ν_1 , and consequently k_2 , will be very large, so that (152) becomes

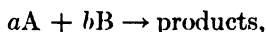
$$\frac{dx}{dt} = \frac{k_1p}{k_2p_1} = k \frac{p}{p_1}, \quad (153)$$

and the rate of reaction should be proportional to the pressure of the reactant, but inversely proportional to that of the strongly adsorbed product. The decomposition of ammonia on platinum at 800° to 1000° c. shows this type of behavior, the reaction being retarded by hydrogen gas firmly held on the surface. An interesting consequence of this marked

retardation is that the time required to reduce the pressure of ammonia to half its initial value is proportional to the pressure; this is a condition for a reaction which is kinetically of zero order. A process of this kind is, however, readily distinguished from those of zero order described above, for in the latter case deliberate addition of the product has little effect on the velocity.

Just as a unimolecular reaction becomes of zero order when one of the products is firmly held on the surface, so a bimolecular process may, for the same reason, prove to be kinetically of the first order. This is the case when nitric oxide decomposes on a surface of heated platinum; the reaction is strongly retarded by oxygen which is one of the products.

Two Reacting Gases.—When a reaction involves two or more gases a number of possibilities may arise according as one or another, or neither, of the reactants or products is strongly adsorbed. It will be assumed, in the first place, that the products are not adsorbed and the surface is sparsely covered with both reactants, so that $1 - \theta$ may be assumed equal to unity. Suppose the reaction is



then according to (141),

$$\alpha_A \mu_A = \nu_A \theta_A \quad \text{and} \quad \alpha_B \mu_B = \nu_B \theta_B \quad (154)$$

for the gases A and B respectively, θ_A and θ_B being the fractions of the surface covered by these molecules. The rate of reaction on the surface is given by

$$\frac{dx}{dt} = k \theta_A^a \theta_B^b, \quad (155)$$

and substituting the values of θ_A and θ_B from (154), it is seen that

$$\frac{dx}{dt} = k_1 p_A^a p_B^b, \quad (156)$$

since μ_A and μ_B are proportional to the respective pressures. The order of the reaction should thus be $a + b$, in agreement with the molecularity.

Although (156) represents the kinetics of the simplest type of heterogeneous reaction involving two gases, there are few instances to which it is clearly applicable. The union of ethylene and hydrogen on a copper surface at 150° to 250° c. is probably one case, and another is the interaction of hydrogen and carbon dioxide, at low pressures of the latter, on platinum at 1000° c. The rate of combination of carbon monoxide and oxygen on crystalline quartz apparently satisfies (156), but it is probable that the reaction is much more complicated than this agreement would imply. On quartz glass the process is undoubtedly inhibited by carbon monoxide, and so the mechanism is not simple. The reaction between hydrogen and oxygen on a gold surface at 120° to 150° c. is similar; the rate is propor-

tional to $p_{H_2}^2 p_{O_2}$, as required by (156), but the mechanism is definitely complex involving strongly adsorbed oxygen.

Retardation by Reactant.—If one of the reacting gases is firmly held on the surface it may have the surprising effect of retarding the reaction as its pressure is increased. Consider a reaction in which one molecule of A and one of B take part. Suppose the gas A is strongly absorbed, and let θ be the fraction of the surface covered by it, where θ is nearly unity; the fraction covered by the gas B is negligible. The corresponding form of (149) is

$$(1 - \theta)\alpha_A\mu_A = \nu_A, \quad (157)$$

the θ on the right-hand side being taken as unity. The rate of reaction may be assumed to be equal to the rate of condensation of the molecules of B on the *free* surface, since there is always sufficient A available in the vicinity to combine with them; hence,

$$\begin{aligned} \frac{dx}{dt} &= k(1 - \theta)\alpha_B\mu_B \\ &= k\alpha_B\mu_B\nu_A/\alpha_A\mu_A = k_1p_B/p_A, \end{aligned} \quad (158)$$

where p_A and p_B are the gas pressures. The rate of reaction should thus be *inversely* proportional to the pressure of the gas strongly held by the surface; the greater the pressure of this gas the slower the reaction.

A number of instances of this type are known: the combination of carbon monoxide and oxygen on platinum at 430° c., at very low pressures, and on quartz glass, is retarded by the former reactant, and the same effect has been observed in the reaction of this gas with nitrous oxide on quartz at 550° c. Although the union of ethylene and hydrogen appears to be a simple process at 150° to 250° c., as seen above, at lower temperatures, viz., 0° to 20° c., ethylene is preferentially adsorbed on copper and the rate of reaction is diminished as its pressure increases. On a nickel surface the hydrogen is most strongly held, and consequently this gas retards the combination with ethylene.

Retardation by Products.—An interesting example of retardation by a product was discovered by M. Bodenstein and C. G. Fink (1907), who observed that in the familiar reaction of sulfur dioxide and oxygen on a finely-divided platinum surface the rate of combination was proportional to the pressure of whichever reacting gas was not present in excess, and inversely proportional to the square root of the pressure of the sulfur trioxide. It is clear that the latter has a retarding influence, and the original explanation offered was that a layer of sulfur trioxide gas was formed through which the reacting molecules had to diffuse before reaching the platinum. If the actual combination occurred rapidly on the surface, in comparison with the rate of diffusion to the surface, then the over-all rate of the reaction would be determined by the time taken for the reactants to diffuse through the sulfur trioxide layer. The thicker the layer the slower would be the rate of diffusion and hence of the chemical

reaction. Calculation has shown, however, that in order to produce the observed effects the sulfur trioxide layer would have to be thick enough to be visible to the eye, and hence an alternative explanation is desirable. In harmony with the views already developed, the retarding effect of sulfur trioxide may be ascribed to its preferential adsorption on the surface. Whichever reactant is present in excess probably occupies most of the surface not covered by the trioxide, and reaction may occur when molecules of the second reactant meet those of the first already on the surface; the rate of reaction will thus be proportional to the pressure of the second reacting gas. If the portion of the surface free from trioxide is inversely proportional to the square root of its pressure, the observed kinetics can be explained.

Adsorption and Desorption as Rate-Determining.—In the heterogeneous processes considered above the rates are determined essentially by the reaction of the gas molecules on the surface of the solid. It is tacitly assumed that the rates of adsorption and desorption of the reactants and products are rapid in comparison with the rate of interaction of the adsorbed molecules. There is a possibility, however, that the slow, rate-determining, process may be the adsorption of the reactants on to the surface, or the desorption of the products. The former is evidently the case in the synthesis of ammonia from nitrogen and hydrogen on a "promoted" (see p. 1152) iron catalytic surface, containing small amounts of aluminum and potassium oxides. The rate of adsorption of nitrogen is probably the slowest stage which determines the over-all rate of the formation of ammonia. In the reverse reaction on the same surface, that is, the decomposition of ammonia into nitrogen and hydrogen, the rate-determining stage will then be the desorption of nitrogen, i.e., its removal from the surface.

It has been found (K. S. Love and P. H. Emmett, 1941) for the latter reaction, which lends itself more readily to experimental study than does the ammonia synthesis, that the rate over a considerable range of pressures is given by

$$-\frac{dp_{\text{NH}_3}}{dt} = k \frac{p_{\text{NH}_3}^{0.60}}{p_{\text{H}_2}^{0.85}},$$

and this apparently curious result can be accounted for in the following manner. If the desorption of nitrogen is rate-determining, the observed reaction rate will presumably be proportional to the rate at which the molecules of this gas leave the surface, and hence to the amount of nitrogen on the surface at any instant. The adsorbed nitrogen molecules will be in equilibrium with the nitrogen in the gas phase over the surface, and if the partial pressure of the latter is represented by p_{N_2} , it may be supposed that the reaction rate will be proportional to $p_{\text{N}_2}^a$, where a is approximately constant (cf. p. 1120). In the gas phase there will also be a state of equilibrium between the ammonia, nitrogen and hydrogen, and if K is the equilibrium constant, it is evident that p_{N_2} is equal to $Kp_{\text{NH}_3}^2/p_{\text{H}_2}^3$. According to the foregoing arguments, the rate of decomposition of ammonia on the surface will be equal to $k(p_{\text{NH}_3}^2/p_{\text{H}_2}^3)^a$, where k is constant for a given temperature. It is now evident that if a is about 0.3, this result becomes almost the same as the observed rate equation quoted above.

From a theoretical and experimental study of the adsorption and desorption of nitrogen on the same promoted iron surface, S. Brunauer, *et al.*, (1942)

were able to derive a value of 0.276 for the exponent a . The close agreement between this and the observed result provides support for the argument that desorption of nitrogen determines the over-all rate of the decomposition of ammonia. It follows from this conclusion that adsorption of nitrogen would be rate-determining in the synthesis of ammonia on the given catalyst. It may be mentioned, too, that the activation energy for the decomposition of ammonia, calculated on the basis of the mechanism just described, is in good agreement with the experimental value.

Adsorption and desorption are probably rate-determining also in the ortho-para hydrogen conversion on nickel and various oxide surfaces, and in the reaction between hydrogen and deuterium on nickel.

Absolute Rate Theory of Heterogeneous Reactions.—The only simple and satisfactory approach to the problem of calculating the rates of heterogeneous reactions, involving gases and a solid surface, has been made by means of the theory of absolute reaction rates. In this connection it is postulated that the activated complex includes a "reaction center," i.e., an atom or molecule of the surface material, in addition to the molecules of reactant. For example, in the bimolecular reaction between the gases A and B, the process would be represented by



where S represents a reaction center, e.g., an atom, on the surface. Assuming ideal behavior the equilibrium between the reactants and the activated state is determined by

$$\begin{aligned} \frac{c_{\ddagger}}{c_A c_B c_S} &= K^{\ddagger} = \frac{Q_{\ddagger}}{Q_A Q_B Q_S} e^{-E/RT}, \\ \therefore c_{\ddagger} &= c_A c_B c_S \frac{Q_{\ddagger}}{Q_A Q_B Q_S} e^{-E/RT}, \end{aligned} \quad (159)$$

where the Q 's are the respective partition function, and E may be taken as the energy which must be acquired by the gaseous reactants before they can form the activated state on the surface; E is therefore the (experimental) activation energy for the heterogeneous reaction. By combining (159) with (105), it follows that

$$\text{Rate of reaction} = c_A c_B c_S \frac{kT}{h} \cdot \frac{Q_{\ddagger}}{Q_A Q_B} e^{-E/RT},$$

and since this can also be represented by $k_r c_A c_B$, where k_r is the specific rate of the reaction for the given surface, it is seen that

$$k_r = c_S \frac{kT}{h} \cdot \frac{Q_{\ddagger}}{Q_A Q_B Q_S} e^{-E/RT}. \quad (160)$$

The partition functions Q_{\ddagger} and Q_S , referring to the activated complex and the surface atoms cannot be evaluated, but it can be shown that their ratio must be approximately unity. Since S is part of the surface

it can have no translational, and probably no rotational, energy; the only contribution to the partition function Q_s is vibrational in character, and this is usually of the order of unity. Similarly, since the activated complex involves S, it will be fixed to the surface, and will have no translational or rotational movement as a whole. The partition function will consist essentially of vibrational and, possibly, internal rotational factors, so that the whole will not differ greatly from unity. No serious error will, therefore, result by taking Q_t/Q_s to be unity, so that (160) becomes

$$k_r = c_s \frac{kT}{h} \cdot \frac{1}{Q_A Q_B} e^{-E/RT}, \quad (161)$$

for a bimolecular reaction. For a unimolecular reaction the result would be the same, but for the omission of one of the Q 's. The partition functions Q_A and Q_B are readily determined from the known properties of the reactants, and there only remains c_s to be assessed in order that the frequency factor A , in the expression $k_r = Ae^{-E/RT}$ for a surface reaction, may be evaluated. For a sparsely covered surface c_s may, in the ideal case, be taken as equal to the number of atoms, e.g., of platinum or other metal, per sq. cm. of surface. If the surface is appreciably covered with one of the reactants or a product, (161) must be modified, and by making the appropriate changes it has been possible to derive a series of equations which not only include the kinetic expressions given above, e.g., (152) and (158), but actually permit the frequency factors to be evaluated. The results obtained in a number of simple cases have been found to be in excellent agreement with experiment.²¹

CATALYSIS

In the year 1835 J. J. Berzelius reviewed a number of apparently diverse observations, made during the previous thirty or forty years, having a factor in common; in every case the nature of the reaction was influenced by the presence of a substance which was itself unchanged in the process. Among the examples quoted were the conversion of starch into sugar by means of acids, the decomposition of hydrogen peroxide in alkaline solution, especially in the presence of certain metals, and the combination of hydrogen and oxygen on spongy platinum. Berzelius was of the opinion that in these reactions a "catalytic" force was manifest, and reaction brought about under the influence of this force he called **catalysis** (Greek: *loosen*). Although the concept of a catalytic force has now been discarded, the term catalysis is used to describe all processes in which the rate of a reaction is influenced by a substance called a **catalyst** that remains chemically unaffected. If the reaction is hindered by the added substance it is said to be a **negative catalyst**, and the word catalyst, when used alone, is almost invariably taken to imply acceleration of the chemical process. In spite of the fact that catalytic changes are of different types,

both homogeneous and heterogeneous, there are certain characteristics common to them all.

Criteria of Catalysis.—(i) *The catalyst is unchanged chemically at the end of the reaction.* It is necessary to emphasize that the amount of the catalyst should be unaltered at the end of the process; consequently it must be unchanged chemically, although it frequently undergoes a physical change. For example, the manganese dioxide employed to catalyze the decomposition of potassium chlorate changes from large crystals to a fine powder (W. H. Sodeau, 1900), and the platinum gauze catalyst used in the oxidation of ammonia shows considerable roughening after some weeks. Since the catalyst undergoes a physical alteration it is evidently involved directly in the mechanism of the reaction, being regenerated, in a somewhat different form, at the end.

(ii) *A small amount of catalyst is often sufficient to bring about a considerable extent of reaction.* Since the catalyst is not used up in the course of the reaction, a small amount will often cause large quantities of material to react. Some catalysts are particularly effective for certain reactions. For example, cupric ions at a concentration of 1 g. ion in 10^9 liters accelerate appreciably the oxidation of sodium sulfite solution by oxygen (A. Titoff, 1903); similarly, colloidal platinum, present to the extent of only 1 g. atom in nearly 10^8 liters, is able to catalyze the decomposition of hydrogen peroxide. Other instances of acceleration by minute traces are known and are utilized to some extent in analytical chemistry. As a general rule the *rate* of a catalytic reaction is proportional to the concentration of the catalyst; this is strictly true only if the process is a homogeneous one and no chain mechanism is involved. It may therefore be regarded as a rough approximation to state that in such cases the velocity increases with the amount of catalyst employed. For many heterogeneous reactions involving gases on the surface of a solid this is also true to some extent, the rate being related to the effective area of the catalytic surface; this explains why finely divided materials are generally the best catalysts.

(iii) *The catalyst does not affect the position of equilibrium in a reversible reaction.* If the catalyst is quite unchanged at the end of the reaction, it can obviously contribute no energy to the system; consequently the second law of thermodynamics requires that the same position of equilibrium should be attained ultimately whether a catalyst is employed or not. Experimental verification has been obtained in connection with the decomposition of gaseous hydrogen iodide and hydrogen selenide, with solid catalysts, and for the transformation of acetone to diacetone alcohol in solution in the presence of various bases. Further, the same equilibrium was reached in the reaction between sulfur dioxide and oxygen using three different catalysts, viz., platinum, ferric oxide, and vanadium pentoxide.

Some instances have been reported of apparent changes in the equilibrium constant in a system as a result of the presence of a catalyst, but these can be readily explained. For example, W. J. Jones and A. Lapworth (1911) found

in the hydrolysis of ethyl acetate, the equilibrium "constant," given by

$$K = C_{\text{Ester}} \times C_{\text{Water}} / C_{\text{Alcohol}} \times C_{\text{Acid}},$$

increased as the proportion of catalyst, hydrochloric acid, was increased; values between 4.4 and 8.9 were obtained as the molecular ratio of water to hydrochloric acid was varied from 381 to 4.6. The assumption made in calculating the equilibrium "constant" was that the concentrations, expressed in moles per unit volume, were a true measure of the activities of the various components. Since ethyl acetate and alcohol undoubtedly form complexes with hydrogen chloride in solution (cf. J. Kendall, 1914 *et seq.*), this is far from being true, apart from other factors causing deviations from ideal behavior. In a case of this kind it is not strictly true that the catalyst remains unchanged; it could no doubt be recovered completely, but the state of equilibrium of the system would then alter correspondingly. The effect of the catalyst in the hydrolysis of ethyl acetate may be considered as a type of solvent influence (cf. p. 1113); the variation in equilibrium constant is then readily explained by changes in the activities of the substances present in the system.

An important consequence of the fact that a catalyst does not in general alter the position of equilibrium is that for a reversible process it must affect the direct and reverse reactions to the same extent, since $K = k/k'$ (p. 1073); thus a good catalyst for the forward reaction will also accelerate the reverse change. In illustration may be quoted the fact that the same catalyst, i.e., hydrochloric acid, can be employed to expedite the esterification of an acid and the hydrolysis of the corresponding ester. Further, sulfur trioxide vapor decomposes more readily in the presence of platinum than of porcelain; the former is also the better catalyst for the combination of sulfur dioxide and oxygen. Many enzymes, which also fall into the category of catalysts, are known to accelerate both forward and reverse reactions.

Catalysts and Initiation of Reactions.—The problem of whether a catalyst can initiate a reaction, or whether it can only accelerate a reaction already taking place, even if extremely slowly, has been a matter of controversy, and it is doubtful if it can yet be considered as solved. W. Ostwald (1895) was the chief exponent of the view that a catalyst cannot start a reaction, but H. E. Armstrong (1885–1903) and, in more recent years, T. M. Lowry (1925–26) have suggested that there are certain reactions which can occur only if a catalyst is present and will not take place in its complete absence. In particular Lowry has cited the mutarotation of glucose as an example of a change which requires the presence of both a proton donor, i.e., an acid, and a proton acceptor, i.e., a base, before it can occur at all; if the proton donor and acceptor are absent there *appears* to be no mechanism whereby the isomeric change from α - to β -glucose can occur. It is not certain, however, that this contention can be completely upheld. It has been found that in chloroform solution no mutarotation of tetramethylglucose was observed for a period of nine days; this does not prove that the change did not occur at all, but that if it did

take place it was so slow as not to be measurable in that time. The combination of hydrogen and oxygen in the presence of finely divided platinum takes place at a measurable rate at ordinary temperatures, but in the absence of the catalyst the reaction is so slow that no appreciable amount of water is formed in many years. The failure to detect water does not prove that no reaction is occurring; the homogeneous, uncatalyzed combination may be so slow that millions of years would be required for a detectable amount of water to be formed. The difference between the two points of view is probably academic; for any reaction to occur it means that the final state of the system must have a lower free energy than the initial state, and if this condition is satisfied there must be some tendency for the change to take place. The rate at which it does so will depend mainly on the necessary energy of activation, which will vary with the mechanism whereby the change is brought about. If in the absence of a catalyst the only available mechanism is one with a high energy of activation, the rate of reaction must inevitably be extremely slow; the change may nevertheless be taking place. A catalyst may, in fact, be defined as a substance that lowers the activation energy of any change that can occur with a diminution of free energy. According to the theory of absolute reaction rates, however, it is more correct to say that a catalyst lowers the *free energy of activation* [cf. equation (118)]; in certain catalytic reactions there is actually an increase in ΔH^\ddagger , but this is more than compensated for by the accompanying increase in $T\Delta S^\ddagger$.

Homogeneous Catalysis: Gases.—For convenience homogeneous and heterogeneous reactions will be considered separately, and a further subdivision will be made according as the reactants are chiefly gaseous or liquid. The type of catalyzed homogeneous reaction involving gases most often quoted is that in which traces of water are supposed to play an important part; it should be pointed out, however, that in recent years some doubt has been cast on the accuracy of the results previously reported. For example, it has been stated that in the absence of water vapor the photochemical combination of hydrogen and chlorine is inhibited; it is now generally agreed, however, that the homogeneous reaction between these two gases under the influence of light is not appreciably affected by moisture. The vapors of mercurous chloride and of ammonium chloride have been said to be undissociated when dry; the latest work on ammonium chloride (H. Braune and S. Knoke, 1928; W. H. Rodebush and J. C. Michalek, 1929) and on mercurous chloride (F. T. Gucker and R. H. Munch, 1937) indicates that dissociation does occur. The results of H. B. Baker (1894) are believed to be due to the slow rate of vaporization of the dried substances, so that the vapor density obtained by the Victor Meyer method would be too high. Since the removal of water does not prevent the dissociation of ammonium chloride, the reported observation, that thoroughly dried ammonia and hydrogen chloride gases do not react, may be questioned, since a catalyst must influence both direct and reverse reactions in a similar manner. Baker

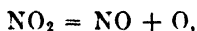
also stated that pure dry nitric oxide and oxygen do not combine, but this has been denied; the dried gases are said to react as readily as in the presence of water vapor (E. Briner, 1926).

In spite of the examples recorded above there are definite instances of the catalytic activity of water in gas reactions. The combination of carbon monoxide and oxygen was believed to be inhibited by drying (H. B. Dixon, 1880), but later work has shown that a reaction does occur in the dried gases, although it involves a mechanism different from that in the moist gases. Dixon suggested that in the presence of water the reaction took place in two stages, viz.,

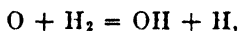


so that the water was regenerated and was thus able to bring about further combination. It is now known, however, that the carbon monoxide-oxygen reaction is a very complex one involving a chain mechanism. Upper and lower pressure limits within which explosion occurs have been observed (cf. p. 1084) at temperatures of about 600° c., and water vapor has little effect on their position; outside these pressure limits the reaction in the dried gases is slow, and the acceleration brought about at higher pressures by hydrogen, water, or other hydrogen-containing substances is very marked. In the absence of a catalyst reaction chains probably start at the walls of the vessel, with oxygen as the chief carrier. When hydrogen or water is present, on the other hand, the process is quite different; chains appear to be initiated in the gas phase and possible carriers are hydrogen atoms and hydroxyl radicals. The kinetics of the reactions are such as to require very complicated mechanisms, but as the subject is far from simple, and has by no means reached finality, further details will be omitted.

An unusual type of catalysis was discovered by H. B. Dixon and investigated by C. N. Hinshelwood (1929). Above 400° c. no appreciably homogeneous combination of hydrogen and oxygen can be observed; if nitrogen dioxide is admitted there is at first no effect, but at a certain pressure, e.g., less than 0.05 mm. when the total gas pressure is 150 mm. of mercury, immediate explosion occurs. As the amount of dioxide is increased an equally sharp upper limit of pressure is found above which only slow reaction between hydrogen and oxygen occurs. These results clearly indicate a branched chain mechanism with the nitrogen dioxide acting as the means whereby the chains are initiated, but the dioxide must also be able to break the chains to account for the inhibition of the explosive reaction as its pressure is increased. The simplest explanation is that the dioxide first decomposes to give atomic oxygen, thus

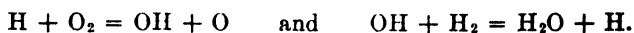


and this reacts with molecular hydrogen,

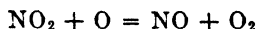


to give hydrogen atoms and hydroxyl radicals, both of which can set off chains,

viz.,



On the other hand the reaction



would terminate the chain, and the probability of this process occurring will increase as the pressure of the dioxide is raised. Although the suggested mechanism does not account for all the facts,* it may be taken as an indication of the method whereby nitrogen dioxide, within definite pressure limits, sensitizes the explosive reaction between hydrogen and oxygen.²²

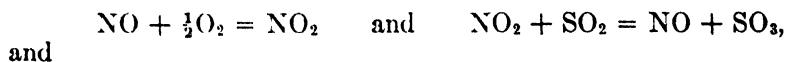
Iodine has been found to exert a striking catalytic effect on the decomposition of the vapors of many ethers and aldehydes; the catalysis is not abnormal, for the rate of reaction is approximately proportional to the concentration of iodine in the vapor. The acceleration of the decomposition may be partly due to the fact that a unimolecular reaction with a high energy of activation is converted into a simple bimolecular process, requiring a collision between the reactant and an iodine molecule and having a low activation energy. The data in Table 186 (C. N. Hinshel-

TABLE 186. ACTIVATION ENERGIES IN THE DECOMPOSITION OF ETHERS

Substance Decomposed	Unimolecular Uncatalyzed	Bimolecular Catalyzed
Methyl ethyl ether	54,500 cal.	38,000 cal.
Diethyl ether	53,000	34,300
Disopropyl ether	61,000	28,500

wood, *et al.*, 1929-30) show the difference between the iodine-catalyzed and uncatalyzed decomposition of some ethers. The striking decrease in the energy of activation and the corresponding increase in the $e^{-E/RT}$ term account for the greater velocity of the catalyzed reactions.

Two other examples of homogeneous catalysis may be mentioned: there is the well-known action of nitric oxide, used in the lead-chamber process, in bringing about the combination of sulfur dioxide and oxygen, and the same substance is able to catalyze the union of carbon monoxide and oxygen. The reactions may be written



and



the nitric oxide being regenerated in each case. These would be examples of what is known as the **intermediate compound theory** of catalysis, but the actual reactions are probably not as simple as the equations would imply. In any case, the presence of the catalyst permits the reaction to take place by an alternative path involving one or more stages which,

* An alternative mechanism, based on the intermediate formation of NO_3 , has been proposed.

requiring small activation energy, take place more readily than the direct combination.

Homogeneous Catalysis in the Liquid Phase.—The most important examples of homogeneous catalysis in the liquid phase are those generally included under the heading of **acid-base catalysis**. It has been long known that acids are able to accelerate the inversion of sucrose and the hydrolysis of esters; these reactions, employing different acids as catalysts, were studied by W. Ostwald (1883–84). After the development of the theory of electrolytic dissociation, S. Arrhenius (1889) showed that the rates of catalytic processes were closely proportional to the concentrations of hydrogen ions, and the conclusion was reached that these were the effective catalytic agents. In Table 187 are given the observed specific

TABLE 187. INVERSION OF SUCROSE IN FORMIC ACID-SODIUM FORMATE SOLUTIONS

Solution	k (obs.)	k (calc.)
0.25 N formic acid	2.55	2.54
+0.025 N formate	0.72	0.75
+0.10 “	0.24	0.24
+0.25 “	0.118	0.117

rates for the inversion of sucrose at 54° c. in aqueous formic acid solution and in mixtures of the acid and sodium formate, which may be compared with the values calculated on the assumption that the velocity is proportional to the concentration of hydrogen ions; the latter is determined from the equation (253) on p. 1003. Other reactions catalyzed by hydrogen ions are the hydrolysis of acetamide and of acetal, and the decomposition of diazoacetic ester by water. On the other hand, hydroxyl ions are particularly effective in the mutarotation of glucose in aqueous solution, in the decomposition of nitrosotriacetoneamine and in the conversion of acetone into diacetone alcohol, and *vice versa*.

In the course of his work S. Arrhenius (1899) noted that the addition of neutral salts to acid solutions often had an unexpected influence in making the catalytic activity greater than that based on the assumption that the hydrogen ion was the only catalyst. Many observations of a similar nature were made subsequently by other workers in connection with a variety of reactions; the effect was found both when the salt had an ion in common with the catalyzing acid and when it had not. For example, the addition of N/8 potassium chloride increased by over 12 per cent the specific rate for the inversion of sucrose by N/4 acetic acid, and in the presence of a mixture of one part of picric acid to ten parts of a picrate, sufficient to reduce very considerably the hydrogen ion concentration, the catalytic activity for the decomposition of diazoacetic ester in alcohol was still one-quarter of that with the same amount of acid alone. Results of this kind led G. Senter (1907), S. F. Acree (1907), H. S. Snethlage (1912), H. S. Taylor (1914) and others to propose what was called by H. M. Dawson (1915) the **dual theory of catalysis in solution**,

which postulated that both hydrogen ions and the undissociated molecules of an acid were able to act catalytically. This view has now been extended, mainly as the result of the work of J. N. Brønsted, of H. M. Dawson, and of T. M. Lowry and their collaborators, from 1923 onward, and it has become evident that not only hydrogen and hydroxyl ions, but also undissociated molecules of acids and bases, and even cations of weak bases and anions of weak acids can have catalytic activity. In fact when a reaction is catalyzed by hydrogen ions all acids, in the widest sense of the term (p. 975), are often found to act as catalysts; this is known as **general acid catalysis**. Similarly **general base catalysis** has been shown to occur with reactions primarily catalyzed by hydroxyl ions; for some processes both acids and bases appear to be effective.

The hydrolysis of ortho-esters, e.g., ethyl orthocarbonate, is catalyzed by all acids, hydrogen ions being the most active, and not by bases; on the other hand, the decomposition of nitramide into nitrous oxide and water, the conversion of diacetone alcohol into acetone, and the decomposition of nitrosotriacetoneamine are catalyzed only by bases, e.g., hydroxyl ions, anions of weak acids, and undissociated molecules of weak bases. The influence of the hydroxyl ions is, however, greatly in excess of that of other catalytic species. The hydrolysis of acetals is catalyzed by hydrogen ions, and these appear to be so much more effective than other acids that "general" acid catalysis is absent; this constitutes an example of **specific catalysis**.

In the hydrolysis of simple esters, e.g., ethyl acetate, in the enolization of acetone, as determined by the reaction with iodine (cf. p. 1135), and in the mutarotation of glucose general acid *and* base catalysis has been observed. The results in Table 188 give some of the catalytic

TABLE 188. CATALYTIC COEFFICIENTS FOR THE MUTAROTATION OF GLUCOSE AT 18°

Acid	<i>k</i>	Base	<i>k</i>
H ₃ O ⁺	1.4×10^{-1}	H ₂ O	9.5×10^{-6}
H ₂ O	9.5×10^{-5}	OH ⁻	6×10^3
Acetic acid	2×10^{-3}	Acetate ion	2.7×10^{-2}
Mandelic acid	6×10^{-3}	Mandelate ion	1.1×10^{-2}

coefficients for the mutarotation of glucose. The hydroxyl ion is clearly the most effective catalyst, then comes the hydrogen ion (H₃O⁺); all acids and bases, however, have a catalytic influence depending on their strengths. Ammonium ions and various metal-ammine ions, which are acidic, are catalytically active, and so also are a number of hydroxo-ions, e.g. [Co(NH₃)₅(OH)]⁺⁺, which have basic properties. Although the existence of general acid and base catalysis accounts for the abnormal results obtained with mixtures of acids and their salts, it does not account for the influence of neutral salts not having an ion in common with the acid. These can, however, be explained by the influence of electrolytes on activity coefficients, and the consequent change in the "kinetic factor" (p. 1115) and in the actual concentration of the ionic catalyst (see p. 1139).

Determination of Catalytic Coefficients.—The over-all velocity of a catalyzed reaction in which a number of catalysts take part is equal to the sum of the rates of the various simultaneous changes involving the different catalysts (p. 1087). If the concentrations of the catalysts are assumed to remain constant, as would be the case for dilute solutions where salt effects are negligible, it follows from (89) and (89a) that the observed specific rate is equal to the sum of the kc terms for the various catalysts, c being the concentration and k the catalytic coefficient. For a reaction undergoing general acid and base catalysis in a solution of a weak acid and its salt, the possible catalysts are hydrogen ions (H_3O^+), acid anions (A^-), hydroxyl ions, molecules of undissociated acid (HA) and water; hence,*

$$k = k_{\text{H}^+\text{CH}^+} + k_{\text{A}^-\text{CA}^-} + k_{\text{OH}^-\text{COH}^-} + k_{\text{HACHA}} + k_{\text{H}_2\text{OCH}_2\text{O}}. \quad (162)$$

In a fairly strongly acid solution the hydroxyl ion concentration is small and the catalytic effect of these ions, as well as that of the water molecules, may be neglected, giving

$$k = k_{\text{H}^+\text{CH}^+} + k_{\text{A}^-\text{CA}^-} + k_{\text{HACHA}}. \quad (163)$$

As the concentration of salt is increased, the amount of acid remaining constant, the hydrogen ion concentration decreases, and the term $k_{\text{H}^+\text{CH}^+}$ may be omitted in comparison with the others, so that at relatively high salt concentration,

$$k = k_{\text{A}^-\text{CA}^-} + k_{\text{HACHA}}, \quad (164)$$

where c_{A^-} may be taken as equal to the concentration of the salt, assumed to be completely ionized, and c_{HA} is that of the weak acid, since it is almost completely nonionized. If the specific rate k is measured in a series of solutions containing a constant known amount of acid, i.e., c_{HA} is constant, but varying amounts of its salt, the plot of k against c_{A^-} should, according to (164), be a straight line at high salt concentrations. The slope of this line will be equal to k_{A^-} and its intercept on the k -axis, when extrapolated to $c_{\text{A}^-} = 0$, should give k_{HACHA} ; from the experimental data, therefore, both k_{A^-} and k_{HA} can be evaluated.

To determine k_{H^+} it is necessary to carry out experiments in a dilute solution, e.g., 0.1 N, of a strong acid, such as hydrochloric acid; the hydrogen ion is then the only effective catalyst, since the chloride ion has a negligible influence, and the concentration of undissociated molecules is very small, i.e., $k = k_{\text{H}^+\text{CH}^+}$. The measured specific rate is then proportional to k_{H^+} and to the concentration of acid, assumed to be completely ionized. For some reactions, e.g., the mutarotation of glucose, k_{OH^-} can be determined in an exactly analogous manner, by measurements of the reaction velocity in the presence of a dilute solution of a strong

* As in Chapter XII the symbol H^+ is used for hydrogen ions, although H_3O^+ should be understood.

base. In other instances, however, this is not possible; for example, in the enolization of acetone the iodine employed to indicate the extent of the change would react with the hydroxyl ions, and an alternative procedure must be adopted.

For a series of solutions having a definite ratio of acid to salt, the hydrogen and hydroxyl ion concentrations may be assumed to be almost constant, so that the catalytic effect of these ions together with that of the water molecules may be included in a constant k_0 , that is,

$$k_0 = k_{H^+}c_{H^+} + k_{OH^-}c_{OH^-} + k_{H_2O}c_{H_2O}, \quad (165)$$

and (162) becomes

$$k = k_0 + k_A c_A^- + k_{HA} c_{HA}. \quad (166)$$

Since c_{HA}/c_A^- is kept constant throughout, it may be put equal to q , and hence

$$k = k_0 + (k_A^- + qk_{HA})c_A^-, \quad (167)$$

so that the plot of k against the salt concentration c_A^- should be a straight line. From the slopes of two such lines, for different ratios of acid to salt, k_A^- and k_{HA} can again be evaluated. The intercept on the k axis, for c_A^- equal to zero, will be equal to k_0 for the hydrogen and hydroxyl ion concentrations in the given acid-salt mixture. By means of a series of mixtures of acid and salt, the value of k_0 for various hydrogen ion concentrations can thus be determined, and the plot of k_0 against $\log c_{H^+}$ is found to be a catenary curve with a definite minimum for k_0 ; this result is to be expected from (165), since c_{OH^-} must increase as c_{H^+} decreases. The condition of the minimum is obtained as follows: in dilute solution the product $c_{H^+} \times c_{OH^-}$ may be taken as constant and equal to k_w (p. 982), when (165) becomes

$$k_0 = k_{H^+}c_{H^+} + k_{OH^-}k_w/c_{H^+} + k_{H_2O}c_{H_2O}. \quad (168)$$

Differentiation with respect to c_{H^+} and equating to zero, gives the condition for the minimum

$$k_{H^+} - k_{OH^-}k_w/(c_{H^+})_{\min.} = 0, \quad (169)$$

and substituting the value for $k_{OH^-}k_w$ so obtained in (168) gives

$$(k_0)_{\min.} = 2k_{H^+}(c_{H^+})_{\min.} + k_{H_2O}c_{H_2O}, \quad (170)$$

where $(k_0)_{\min.}$ is the minimum value of k_0 , obtained at the hydrogen ion concentration $(c_{H^+})_{\min.}$. Since k_{H^+} is known from separate experiments k_{H_2O} can be calculated, the concentration of water c_{H_2O} being taken as $1000/18.02 = 55.5$ mole per liter.* Now that k_{H^+} and k_{H_2O} are available, k_{OH^-} can be determined by substitution in (168), using the value of k_0 at any definite hydrogen ion concentration. From (165) and (170) it can be

* The quantity $k_{H_2O}c_{H_2O}$ is often left in this form, when it is called the specific rate of the "spontaneous reaction," since it is the contribution of the solvent only.

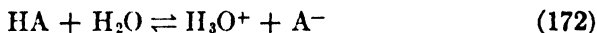
readily shown that when k_0 is a minimum the catalytic contributions of hydrogen and hydroxyl ions are equal; that is, $k_{H^+}c_{H^+} = k_{OH^-}c_{OH^-}$.

When the reaction shows only acid catalysis the experimental determination of the catalytic coefficients can be simplified. The recommended procedure is to make use of (167) for constant ratio mixtures of acid and salt; since there is no basic catalysis k_{A^-} is now zero, and this equation becomes

$$k = k_0 + qk_{HAC_A^-}, \quad (171)$$

so that k_0 can be obtained by extrapolation of the linear plot of k against c_{A^-} , the slope giving k_{HA} . If k_{H^+} is known from experiments in acid solution k_{H_2O} can be derived from (165), since k_{OH^-} is zero. If bases only act as catalysts, the coefficients required are k_{OH^-} , k_B and k_{H_2O} , where k_B is that for the undissociated base; they are best evaluated from measurements in solutions of a weak base and its salt. The appropriate equations can be readily developed by analogy with those given above for acid-salt mixtures.

Catalytic Coefficient and Dissociation Constant.—Since all acids have in common the tendency to lose a proton (p. 975), it is reasonable to suppose that their catalytic action is dependent on the magnitude of this tendency which should be related at least approximately, to the dissociation constant of the acid. Consider the acid HA in aqueous solution: the equilibrium



is set up, and the dissociation constant K_a of the acid, which is the equilibrium constant of the system (cf. p. 977), is equal to the ratio of the specific rates of direct and reverse reactions, i.e.,

$$K_a = \frac{k}{k'} = \frac{\text{Specific rate of reaction of } HA + H_2O}{\text{Specific rate of reaction of } H_3O^+ + A^-}. \quad (173)$$

The quantity k is a measure of the tendency of the acid HA to lose its proton to a water molecule and hence it might be expected to be related to the catalytic effect k_{HA} of the acid. In practice, however, it is not possible to separate K_a into k and k' , and J. N. Brønsted (1924) suggested the approximate relationship

$$k_{HA} = GK_a^x, \quad (174)$$

where G and x are constants for a given reaction, and k_{HA} is the catalytic coefficient of the acid having a dissociation constant K_a . According to (174) the stronger an acid the greater will be its catalytic activity, and a similar relationship may be considered to apply to basic catalysts. Taking logarithms (174) becomes

$$\log k_{HA} = \log G + x \log K_a, \quad (175)$$

so that when the logarithms of the catalytic coefficients of various acids are plotted against the logarithms of the respective dissociation constants

a straight line should result; this has been found to be approximately true for a number of reactions. The data in Fig. 251 are for the mutarotation of glucose (J. N. Brønsted and E. A. Guggenheim, 1927). It may be noted that the values of $k_{\text{H}_3\text{O}^+}$, $k_{\text{H}_2\text{O}}$ and k_{OH^-} , which are the catalytic constants of the two acids H_3O^+ and H_2O , and of the base OH^- , respectively, often fall on the appropriate curves, provided the correct dissociation constants are employed.

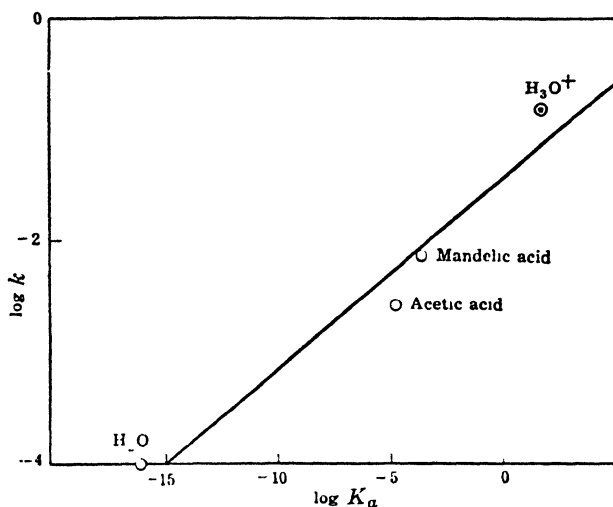


FIG. 251. Acid catalyzed mutarotation of glucose

The fact that the catalytic effects of acids, varying in strength from the very strong acid H_3O^+ to the very weak acid H_2O , can be brought into one scheme lends support to the suggestion that it is a common property, i.e., the tendency to lose a proton, which determines their catalytic activity. Further confirmation of this point of view is the discovery that acids are able to act as catalysts in nonionizing solvents, e.g., benzene, where hydrogen ions do not exist. J. N. Brønsted and R. P. Bell (1931) found that the reaction between diazoacetic ester and phenol in benzene solution was catalyzed by acids, and that the coefficients are related to the dissociation constants in *water* by a relationship of the type of (174). Similar results have been obtained for the rearrangement of N-bromoacetanilide and several other reactions in various nondissociating solvents (R. P. Bell, 1934 *et seq.*).

Mechanism of Acid-Base Catalysis.—Since acid catalysis depends on the tendency of the acid to lose a proton, and base catalysis depends on the tendency for the base to gain a proton, it appears that the addition and removal of protons must be an essential part of the mechanism of the catalyzed reactions under consideration. The mutarotation of glucose, for example, probably involves the addition of a proton to the molecule

of α -glucose at one point, followed, or accompanied, by its removal at another point; thus if GH represents α -glucose and HG is the isomeric β -glucose, the catalyzed change is represented by

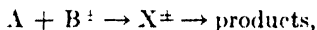


the first proton is supplied by an acid, whereas the one discarded is removed by a base. According to this view the mutarotation process requires the presence of an acid and a base, in harmony with the observations that, in the absence of water, the base pyridine is a poor catalyst for the mutarotation of tetramethylglucose, and so also is the acid *m*-cresol, but a mixture of the two is a powerful catalyst (T. M. Lowry, 1925). The other acid-base catalyzed isomeric change, viz., the enolization of acetone, probably takes place in a similar manner; thus,



In a reaction, such as the decomposition of nitramide, which is catalyzed almost exclusively by bases, the mechanism must be somewhat different; the same applies to the hydrolysis of acetals, where only acids are catalytically active. It is probable that addition and removal of a proton, or *vice versa*, occurs in two stages, but that only one of them can be rate-determining, the other occurring very rapidly in comparison.

Salt Effects.—The influence of salts in catalysis by acids and bases may be divided into two categories, viz., primary and secondary kinetic salt effects; the former has already been considered in connection with ionic reactions in general (p. 1115), and its influence in ionic catalysis will now be examined. Many catalytic reactions, such as those described above, involve a neutral molecule, e.g., sucrose, acetone, or an ester, and a univalent ion, e.g., hydrogen or hydroxyl, so that the formation of the activated complex may be written



where A is the neutral molecule, and B^\pm is a univalent positive or negative ion. According to (140) the specific rate should be independent of the ionic strength, since $z_A z_B$ is zero, provided the solutions are dilute enough for the limiting law of Debye and Huckel to be approximately applicable; this has been found to be true for a number of reactions. At appreciable concentrations, however, it is necessary to use the extended relationship (p. 968) of the form

$$-\ln f = \alpha \sqrt{\mu} - \beta \mu, \quad (176)$$

and the activity coefficient of a neutral molecule will also vary with the ionic strength in accordance with the equation $\ln f = \beta \mu$. For a reaction involving a neutral molecule and a univalent ion, the logarithm of the kinetic activity factor is given by

$$\ln \frac{f_A f_B}{f_\ddagger} = \mu(\beta_A + \beta_B - \beta_\ddagger) = \mu\beta_s; \quad (177)$$

$$\therefore \frac{f_A f_B}{f_\ddagger} = e^{\mu\beta_s}, \quad (178)$$

where $\beta_s = \beta_A + \beta_B - \beta_I$. Substituting in (133), it follows that

$$k = k_0 e^{u\beta_s}. \quad (179)$$

If β_s is small, as it is generally, it is possible to write (179) as

$$k \approx k_0(1 + u\beta_s). \quad (180)$$

As a result of the primary salt effect, therefore, the observed specific rate should be a linear function of the ionic strength, since k_0 and β_s are constants for the given process. This has been verified for several reactions, e.g., inversion of sucrose, hydrolysis of esters and acetals, and the decomposition of nitramide. The actual variation of k with increasing ionic strength depends on the magnitude and sign of β_s , and this, like the individual β values, is determined by *all* the ions present in the solution, not merely by those acting as catalyst. In practice β_s is found to be relatively large when hydrogen and hydroxyl ions are involved, but it is smaller for other ions.

Although the foregoing treatment has been found to be applicable in a number of instances, cases of its failure have been reported. This may be due to fundamental defects in the argument, but it is more likely to be caused by complications, such as ion-dipole interactions between the added ions and the reacting molecule and similar interactions involving other constituents of the system, which are not taken into account in deriving the conclusions stated above.

Secondary Kinetic Salt Effect.—The secondary effect operates by changing the actual concentrations of the catalytically active ions; this is important when the latter are produced by the dissociation of a weak electrolyte, but it is negligible when a strong acid or base is the catalyst. Suppose a mixture of a weak acid and its salt, at definite concentrations, provides hydrogen ions which are the effective catalyst; then since (cf. p. 980)

$$K_a = \frac{a_{H_3O^+} \times a_{A^-}}{a_{HA}} = \frac{c_{H_3O^+} \times c_{A^-}}{c_{HA}} \cdot \frac{f_{H_3O^+} \times f_{A^-}}{f_{HA}}, \quad (181)$$

the concentration of hydrogen ions is given by

$$c_{H_3O^+} = K_a \frac{c_{HA}}{c_{A^-}} \cdot \frac{f_{HA}}{f_{H_3O^+} \cdot f_{A^-}}. \quad (182)$$

For a definite ratio of acid to salt the quantity $K_a c_{HA}/c_{A^-}$ is constant; putting this equal to K , then

$$c_{H_3O^+} = K \frac{f_{HA}}{f_{H_3O^+} \cdot f_{A^-}}. \quad (183)$$

Since the activity term varies with the ionic strength of the medium, so also will the hydrogen ion concentration and hence its catalytic influence; this is the secondary kinetic salt effect. The velocity of a reaction involving hydrogen ions as catalyst is proportional to the concentration of these ions, and consequently, if the primary salt effect remains constant,

$$k = k_0 \frac{f_{HA}}{f_{H_3O^+} \cdot f_{A^-}}, \quad (184)$$

where k_0 is a constant, being the value of k when the activity coefficient factor is unity, that is, at infinite dilution. Utilizing the Debye-Huckel limiting law for the variation of activity coefficient with ionic strength in dilute solution, and assuming f_{HA} for the undissociated molecules to be unity, then

$$\log k = \log k_0 + 1.02 \sqrt{\mu} \quad (185)$$

for an aqueous solution at 25° c. For an acid-salt mixture the secondary salt effect will result in an increase in the specific rate with increasing concentration of electrolyte.

If the catalyzing acid happens to be a positively charged ion, e.g., NH_4^+ or $\text{Fe}(\text{H}_2\text{O})_6^{+++}$, the results are different. For the ammonium-ion acid the appropriate form of (184) is

$$k = k_0 \frac{f_{\text{NH}_4^+}}{f_{\text{H}_3\text{O}^+} f_{\text{NH}_3}} \quad (186)$$

Since the two univalent ions H_3O^+ and NH_4^+ will have the same activity coefficients in dilute solution, and that of the undissociated molecules is unity, it follows that the specific rate will be independent of the ionic strength, provided the solution is not too concentrated. If the acid has a triple positive charge, e.g., $\text{Fe}(\text{H}_2\text{O})_6^{+++}$, then

$$k = k_0 \frac{f_1}{f_{\text{H}_3\text{O}^+} f_2}, \quad (187)$$

where f_2 and f_3 are the activity coefficients of bi- and ter-valent ions, respectively; the Debye-Huckel equation then gives

$$\log k = \log k_0 - 2.04 \sqrt{\mu} \quad (188)$$

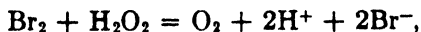
at 25° c., so that the secondary salt effect should be negative. The effects to be expected for other acids and for bases of various types can be derived in an exactly analogous manner.

The existence of the secondary kinetic salt effect has been proved in a number of instances, e.g., the catalytic decomposition of nitrosotriacetoneamine by hydroxyl ions, and the hydrolysis of ethyl acetate with hydrogen ions as catalyst.²⁴

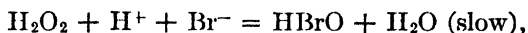
Catalytic Reactions in Solution.—Following acid-base catalysis, the most fully studied case of catalysis in solution is the decomposition of hydrogen peroxide in the presence of various ions; the systems consisting of a halogen and a halide ion are of particular interest. If a halide ion, preferably in acid solution, is added to hydrogen peroxide, a steady state is soon reached in which the halogen and halide ion concentrations remain constant and the net reaction is the decomposition of the peroxide. Careful study of the mechanisms of the processes involved has shown that at the steady state two "compensating reactions" occur at the same rate; for example, when the bromine-bromide ion system is the catalyst the processes are



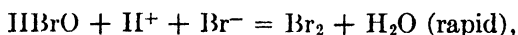
and



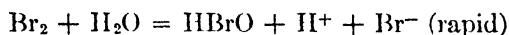
with the result that the peroxide is decomposed catalytically. It is probable that both these reactions occur in stages; they are the same for chlorine but differ slightly when iodine is the catalyst. With bromine the stages are



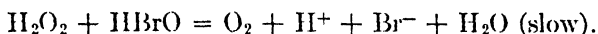
and



the sum being the first of the reactions given above. The stages in the compensating process are

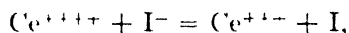


and



When iodine is the catalyst the main reactions involve IO^- ions, instead of molecules of undissociated acid HIO ; otherwise the stages are analogous to those for the other halogens (E. Abel, 1920, 1928-30; W. C. Bray, 1923-30).²⁵

Homogeneous catalytic processes in solution appear to be generally characterized by the formation of an intermediate compound which then decomposes, as in the cases recorded above. Another example is the use of silver ions to catalyze oxidations by persulfates, Ag^{+++} ions being the probable intermediate. The reaction between ceric and thiosulfate ions occurs slowly, but a small amount of iodide acts catalytically and makes the reaction rapid enough to be useful for purposes of volumetric analysis. The slow reaction is replaced by two or more fast stages. The first of these is probably



and this may be followed by formation of iodine molecules which react rapidly with the thiosulfate ions or the latter may react directly with iodine atoms; in any event iodide ions are regenerated in this stage. The iodide ion similarly catalyzes the reaction between arsenious and ceric ions in solution; the velocity is directly proportional to the concentration of iodide ions, so that the rate-determining step is probably the reaction between ceric and iodide ions depicted above. The iodine atoms, or molecules, subsequently react rapidly with the arsenious ions, reforming iodide ions.

Reaction of Organic Compounds.—The use of aluminum chloride as catalyst in the Friedel-Crafts reaction is an example of a homogeneous reaction in the liquid phase, although the catalyst is added in the solid form. It is well known that aluminum chloride forms compounds with acid chlorides, and many workers have held the view that these are intermediates in the reaction between acid chlorides and hydrocarbons. It has been pointed out, however, that formation of the intermediate compound is alone inadequate to account for the Friedel-Crafts reaction, and that there must also be some form of loosening of the $\text{C}-\text{Cl}$ bond (J.

Böeseken, 1907). The suggestion was made that the first stage in the process is the formation of an ionized addition compound, viz., $[\text{CH}_3\text{CO}]^+ [\text{AlCl}_4]^-$, which subsequently reacts with the hydrocarbon (H. Meerwein, 1927), and evidence for this view has been obtained from conductance measurements, and from a study of the Friedel-Crafts reaction with aluminum chloride containing radio-chlorine (cf. p. 175). In the interaction between benzene and acetyl chloride, containing normal chlorine, in the presence of the radioactive chlorine, the activity of the escaping hydrogen chloride was found to be the same as that of the residual solution. It is evident, therefore, that complete interchange occurs between the chlorine atoms in the acid chloride and those in the aluminum salt; such a result can best be explained by postulating the formation of the AlCl_4^- ion as an essential stage in the Friedel-Crafts reaction. Similar results were obtained with *tert.*-butyl chloride in place of acetyl chloride (F. Fairbrother, 1937).

Negative Catalysis and Inhibition.—When a substance is able to diminish the rate of a reaction it is said to be a **negative catalyst** or an **inhibitor**; for example, the oxidation of sodium sulfite solution by oxygen gas is inhibited by small amounts of mannitol, benzyl alcohol, and other primary and secondary alcohols, aniline, benzaldehyde and certain alkaloids, particularly brucine (S. L. Bigelow, 1898). The oxidation of benzaldehyde and of other aldehydes can be inhibited in an analogous manner. The explanation first proposed to account for negative catalysis was that the inhibitor combined with and so removed a positive catalyst already present. A. Titoff (1903) showed, in this connection, that a very minute concentration of cupric ions was able to catalyze the reaction between sulfite solution and oxygen (see p. 1127), and that the effect of the positive catalyst could be counteracted by the addition of a negative catalyst, such as mannitol or cyanide ions. Another possibility is that the walls of the vessel are able to act catalytically and that the inhibitor poisons them (p. 1151). The oxidation of liquid benzaldehyde by oxygen gas, for example, has been said to be a wall reaction; negative catalysts might affect the nature of the wall and so influence the reaction velocity (O. M. Reiff, 1926). The view more widely adopted at present is based on the suggestion of J. A. Christiansen (1924) that the function of a negative catalyst is to prevent the formation or propagation of reaction chains. If the process involves very long chains, a small quantity of inhibitor will be able to reduce the velocity to a marked extent by breaking the chains very soon after they are formed. The work of H. J. Bäckström (1927) has provided evidence for this view. The quantum efficiency (see p. 1160) for the photochemical oxidation of benzaldehyde and enanthaldehyde and of solutions of sodium sulfite by oxygen gas is so high that reaction chains are undoubtedly involved. The addition of small amounts, viz., about 0.01 per cent, of inhibitors, e.g., anthracene or diphenylamine for benzaldehyde; α - or β -naphthols for enanthaldehyde, and various alcohols for sodium sulfite, decreases the quantum

efficiencies, showing that the reaction chains are being broken. Since each chain involves about 10,000 molecules in the case of benzaldehyde and about 50,000 molecules of sodium sulfite, it is easy to understand the enormous influence of relatively small amounts of inhibitors. The inhibitors mentioned above are also able to reduce the rate of the dark, or thermal, reactions so that in these cases they presumably also function as chain breakers. The rate of the thermal reaction is given by the expression

$$\text{Rate of oxidation} = \frac{k}{k_{1c} + k_2}, \quad (189)$$

where k , k_1 and k_2 are constants and c is the concentration of the inhibitor. Comparison of this equation with (83) suggests that chains are broken in two ways, one being by the inhibitor. If the concentration of the latter is appreciable the constant k_2 may be neglected in comparison with k_{1c} , and the rate of reaction should be inversely proportional to the concentration of the negative catalyst, as found in a number of instances. It will be noted that no part has been attributed to the surface of the vessel, but it is not impossible that the reaction chains are initiated there. During the oxidation of sodium sulfite solutions the inhibitors are also oxidized; for example, isopropyl alcohol gives acetone, and benzyl alcohol gives benzaldehyde. This is to be expected if the inhibitors break the chains by themselves combining with the chain carriers, which are probably active oxidizing agents (H. J. Bäckström and H. N. Alyea, 1929). Incidentally this provides an example of the phenomenon of **autoxidation** or **induced oxidation**, first observed by C. F. Schönbein (1858); the alcohols used as inhibitors are not normally oxidized by oxygen gas, but the oxidation occurs only when the sulfite ions are simultaneously oxidized.

Another type of negative catalysis is found in the decomposition of various organic acids, e.g., formic, oxalic, malic and citric acids, with sulfuric acid as catalyst, when small quantities of water, potassium or ammonium sulfates, and various organic oxygen compounds, act as inhibitors. The decomposition of the organic acid presumably involves primary formation of an addition compound between the reactant and sulfuric acid, and so it has been suggested that the negative catalysis is due to preferential compound formation between the sulfuric acid and the inhibitor. It is well known that such compounds exist, at least in solution, because of the basic nature of organic oxygen derivatives. It has been pointed out, however, that if this were the correct explanation the influence of various negative catalysts should fall in the same order no matter which acid is being decomposed; this is not so, and the possibility has been considered that a ternary compound, involving organic acid, sulfuric acid and inhibitor, is formed (E. O. Wiig, 1930).

Negative Catalysis in Gas Reactions.—Negative catalysis in gas reactions is probably also to be ascribed to the breaking of reaction chains. The phenomenon of "knocking" in the internal combustion engine appears to be connected with the propagation of nonstationary branched

chains (p. 1083), and the "anti-knock" materials, e.g., lead tetraethyl or nickel carbonyl, act by producing metallic compounds, e.g., peroxides, which react with the chain carriers and destroy them. Nitric oxide reduces the rate of thermal decomposition of the vapors of acetaldehyde and ethyl ether; it seems to be able to combine with and so remove free radicals. It thus decreases the velocity of a reaction involving chains propagated by such radicals. Propylene has a similar effect in retarding chain reactions (see p. 1082). The halogens, particularly iodine, act as negative catalysts for the homogeneous combination of hydrogen and oxygen; the reaction involves a chain mechanism (p. 1082) and the inhibitors presumably act by destroying the carriers.²⁶

Heterogeneous Catalysis Involving Gases.—Interaction between gases in the presence of a solid catalyst is a common phenomenon; in fact all the heterogeneous reactions, previously considered from the standpoint of their kinetics (p. 1118), are instances. These processes are all catalyzed by the material on the surface of which the reactions occur; in the absence of the solid the velocity is generally very small. It is not impossible that all gas reactions take place to some extent on the surface of the containing vessel although in certain instances the homogeneous process definitely predominates, provided the available surface is not very large. In attempting to understand the phenomena of catalysis it is necessary to consider the actual mechanism of heterogeneous reactions, and also the reasons why the molecules are apparently more reactive on the catalyst surface than in the gas phase.

Even before catalytic phenomena had been correlated by Berzelius, the question of the influence of surfaces in bringing about the combination of gases had been considered by M. Faraday (1833). He concluded that by virtue of the attraction which the solid exerted on the reacting gases, the latter are concentrated near the surface with the result that interaction occurs more readily. In modern terminology this would imply that, as a consequence of adsorption of the gases by the catalyst, a layer is formed in which the reactants are at a higher concentration than in the bulk of the gas, so the rate of reaction would be increased. Somewhat similar views, with slight modification, concerning the relationship between catalysis and adsorption were held for many years; for example, one possibility was that the molecules of one of the reacting gases are adsorbed by the surface in the form of a dense layer, and consequently the probability of impact with the molecules of the other reactant would be greatly increased.

A much clearer outlook has been obtained, however, as the result of the work of I. Langmuir (1916); he suggested that adsorption on a solid surface frequently involves forces similar to those concerned in chemical valence, and that since such forces are exerted over distances of the order of $2-3 \times 10^{-8}$ cm., only a unimolecular layer of adsorbed gas will form. According to this view, the conclusion of earlier workers that there exists near the surface of a solid a relatively thick layer of gas mole-

cules is erroneous. It would seem, therefore, that the adsorbed molecules of reactant are attached to the surface by forces virtually equivalent to chemical bonds. In other words, this means that the surface is to be regarded as one of the reacting species; this is the same as the assumption made in postulating, as on p. 1125, that the activated complex involves the reacting molecules and also an atom (or group of atoms) of the surface material. In a catalyzed reaction between two gases, for example, taking place on the surface of a solid, it is probably necessary that the reacting molecules should first be adsorbed at adjacent points on the surface, in the manner just described. If the molecules possess sufficient energy they may form the appropriate activated complex involving the surface atom (or atoms). The adsorbed activated complex will then decompose at a definite rate, forming the reaction products which are eventually desorbed from the surface and pass into the gas phase. If one of the reactants, or products, is strongly adsorbed, it will prevent access of the other, or both, reactants to the surface, and consequently the reaction will be retarded as shown above. The ideal conditions for catalysis are that both gases should be adsorbed to an appreciable extent, but neither so strongly that it effectively excludes the other.

Some interesting conclusions concerning surface reactions can be drawn from the absolute rate theory equations; the general results are the same irrespective of the molecularity of the reaction, and so a bimolecular process will be considered. For a *homogeneous* gas reaction between A and B, the specific rate is given by (110) as

$$k_{\text{hom.}} = \frac{kT}{h} \cdot \frac{Q_{\ddagger}}{Q_A Q_B} e^{-E_{\text{hom.}}/RT},$$

while for the heterogeneous process, assuming the same molecularity, (161) is applicable, i.e.,

$$k_{\text{het.}} = c_s \frac{kT}{h} \cdot \frac{1}{Q_A Q_B} e^{-F_{\text{het.}}/RT},$$

so that

$$\frac{k_{\text{het.}}}{k_{\text{hom.}}} = \frac{c_s}{Q_{\ddagger}} e^{\Delta F/RT}, \quad (190)$$

where ΔE , equal to $E_{\text{hom.}} - E_{\text{het.}}$, is the difference in the activation energy for the homogeneous and heterogeneous reactions. The partition function, for unit volume, of most molecules lies in the range of 10^{24} to 10^{30} , and so 10^{27} may be taken as an average value for Q_{\ddagger} ; further, for 1 sq. cm. of surface c_s can be assumed to be about 10^{15} , since this is approximately the number of atoms on a smooth surface. If these figures are inserted in (190), it is seen that

$$\frac{k_{\text{het.}}}{k_{\text{hom.}}} \approx 10^{-12} e^{\Delta E/RT},$$

the specific rate for the heterogeneous reaction referring to 1 sq. cm. of (smooth) surface. It is evident, therefore, that if a heterogeneous process is to be as rapid as a homogeneous one involving the same reactants at the same pressures, either a very large surface, e.g., 10^{12} sq. cm., must be employed, or alternatively the activation energy for the surface reaction must be considerably less than that for the homogeneous process. At 500°K. , for example, the difference in the heats of activation, i.e., ΔE , would need to be about 28 kcal. if the rate of the heterogeneous reaction for 1 sq. cm. of surface is to be the same as for the reaction occurring

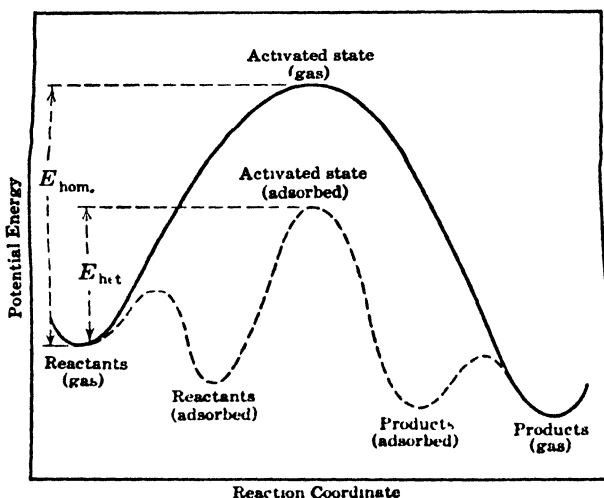


FIG. 252. Potential energy curves for same reaction: homogeneous and heterogeneous

exclusively in the gas phase. In actual catalytic processes the surface exposed to the gases is made as large as possible,* and there is also a marked decrease in the activation energy. This is illustrated by the experimental results quoted in Table 189, for a number of reactions which have been studied both as homogeneous and heterogeneous processes.

TABLE 189. ACTIVATION ENERGIES FOR HOMOGENEOUS AND HETEROGENEOUS REACTIONS

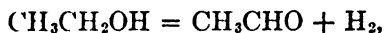
Decomposition of	Surface	E_{het}	E_{hom}
Hydrogen iodide	Gold	25.0 kcal	44.0 kcal
Hydrogen iodide	Platinum	14.0	44.0
Nitrous oxide	Gold	29.0	58.5
Nitrous oxide	Platinum	32.5	58.5
Ammonia	Tungsten	39.0	> 80
Ammonia	Platinum	44.3	> 80

* Adsorption measurements have shown that the actual surface area of 1 cc. of a silica catalyst is about 5×10^6 sq. cm.; for many other catalysts it is of the order of 10^6 to 10^7 sq. cm. (S. Brunauer and P. H. Emmett, 1937).

The higher reaction rates on surfaces of large effective area can be readily understood.

It is instructive, in connection with the fact that the activation energy for the surface reaction is less than for the gas reaction, to examine the potential energy curve in Fig. 252, which is analogous to Fig. 248. The full line depicts the variation of potential energy of the reacting system in the decomposition coordinate for the homogeneous reaction, while the dotted line is for the heterogeneous process. The difference between $E_{\text{hom.}}$ and $E_{\text{het.}}$ is seen to be equal to the difference in potential energy between the activated state in the gas phase and on the surface, and hence is equal to the heat of adsorption of the activated complex. It appears, therefore, that the stronger the bonds between the surface atoms and the reacting molecules in the activated complex, the greater will be the rate of the surface reaction. The probable significance of this conclusion in chemical terms is that formation of relatively strong attachments between parts of the reactants and the surface atoms results in a loosening of the bonds in other parts of the molecules; a smaller amount of energy is then required to form the activated state which is a necessary intermediate in the reaction. This aspect of the subject will be considered further below, from another point of view.

Function of the Surface.—Although one of the functions of a catalytic surface is the general one of providing a large area, there must also be a specific property which determines the adsorption of the activated complex, and the corresponding decrease of activation energy. The nature of the surface is evident in the reaction between ethylene and bromine vapor; at one time this was thought to be a homogeneous process, but it is now known to take place almost entirely on the walls of the reaction vessel. The process is rapid on a glass surface or on one covered with cetyl alcohol; it is still more rapid on stearic acid, but on paraffin wax very little reaction occurs. The different behavior of ethyl alcohol on different catalytic materials is even more striking. If the vapor is passed over a copper or a nickel catalyst at 300° to 400° c., the products are acetaldehyde and hydrogen, i.e., "dehydrogenation" occurs, viz.,



but if the catalyst is alumina at the same temperature, ethylene and water are obtained, i.e., "dehydration" takes place, viz.,



Similar alternative decompositions have been observed with formic acid vapor, the products being carbon dioxide and hydrogen with the former catalysts, and carbon monoxide and water with the latter. On other oxides both types of reaction occur to different extents; thus, on titanium oxide the main process is dehydration, but on zinc oxide dehydrogenation occurs. It is highly significant that the dehydrogenation catalysts, such as copper, nickel and zinc oxide, adsorb hydrogen very strongly, whereas

the dehydration catalysts, e.g., alumina, adsorb water vapor preferentially. The substances of the former type evidently have a marked attachment for atoms of hydrogen, while the latter tend to form bonds with the atoms in water. It may be noted, incidentally, that since a given catalyst should accelerate both direct and reverse reactions,* the dehydrogenation catalysts should also be effective for hydrogenation processes. As is well known, nickel is particularly useful for this purpose.

Non-Uniformity of Surfaces.—General considerations suggest that the surface of a catalyst is not uniform, but the variations actually observed are surprising. Ethylene and hydrogen are both adsorbed to some extent by copper, and on this catalyst they react to form ethane; a trace of mercury reduces the amount of ethylene adsorbed to 80 per cent and that of hydrogen to about 5 per cent of the values on a clean surface, but the catalytic efficiency is at the same time reduced to less than 0.5 per cent of its initial value. This result suggests, in the first place, that the two gases are adsorbed on different parts of the surface and, in the second place, that catalytic action occurs at only a small proportion of the centers of adsorption. The latter conclusion is supported by the observation that traces of carbon monoxide, sufficient to cover only a fraction of the available surface, are able to inhibit completely the reaction between ethylene and hydrogen on copper (R. N. Pease, 1923, 1925).

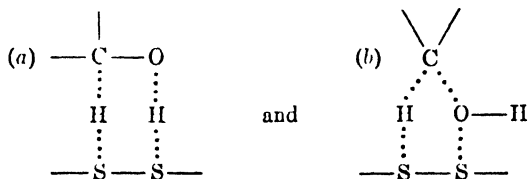
Further striking evidence concerning the nonuniformity of surfaces, and incidentally that ordinary adsorption may not be effective in catalytic action, is shown by a study of the reaction of carbon dioxide and hydrogen on platinum at 1000° c. (C. N. Hinshelwood, 1925). The total adsorption of hydrogen is considerable and that of carbon dioxide is negligible, yet the rate of reaction is proportional to p_{H_2}/p_{CO_2} , indicating (cf. p. 1123) that carbon dioxide is preferentially adsorbed on those parts of the surface where reaction is occurring. Although the reaction between carbon dioxide and hydrogen is thus inhibited, the simultaneous decomposition of nitrous oxide is unaffected; this latter process must, therefore, take place on a different part of the surface from the former. Similarly, the catalytic activity of a platinum wire "poisoned" for the decomposition of ammonia is unchanged for the dissociation of hydrogen iodide, and the parts of a copper catalyst on which combination between ethylene and hydrogen occurs are not those effective for the decomposition of nitrous oxide.

Nature of Active Centers.—Since no solid surface can be ideally smooth, the "residual field" of the atoms responsible for adsorption and catalysis will not be uniform. An atom on a flat surface is attached to similar atoms over a solid angle of 180°, but an atom at the apex of a peak, or at the edge or corner of a crystal, will have a much smaller fraction of its electrostatic attraction shared by others, and so its residual field will be large. It has been suggested (H. S. Taylor, 1925) that such

* Apart from the thermodynamic requirement, this conclusion also follows from Fig. 252.

"unsaturated" atoms generally constitute the **active centers** involved in catalysis. Lines of demarcation or discontinuity, e.g., edges of crystals and grain boundaries, or cracks or imperfections in crystals, all of which represent positions of unsaturation, probably function as active centers. Although the factors determining these centers may be in doubt it is quite certain that they exist on catalytic surfaces, and that they have a definite function in catalysis.

Because of the strong residual field at an active center the conditions are available for the attachment of the activated complex, thus leading to a decrease in activation energy, as explained above. A further development of this concept is the "multiple adsorption" suggestion of R. E. Burk (1926) and A. A. Balandin (1929), and others; it is supposed that the molecule is activated only when it is adsorbed at two or more of the active centers, so that a direct strain is produced in a particular bond or bonds. This hypothesis permits of an interpretation of the alternative modes of decomposition of substances such as alcohols and formic acid vapor according to the spacing of the active centers, and other properties, of the catalytic surface. Two simple types of attachment can be postulated, viz.,



where S represents the active centers, which are not necessarily identical. In case (a), which would represent the behavior on a dehydrogenation catalyst, it is evident that the removal of two atoms of hydrogen from the reactant would be favored; in case (b), the reaction would clearly be dehydration.

Striking evidence for the view that a particular part of an alcohol molecule must be attached to the surface for it to be activated is provided by the work of W. G. Palmer and F. H. Constable (1925-28) on the dehydrogenation of alcohols on a copper catalyst. The rate of decomposition of the primary alcohols, ethyl, propyl, butyl, isobutyl, and isoamyl, on a given catalyst at a definite temperature, is independent of the hydrocarbon chain length. The activation energy of the reaction is also the same for the different alcohols. The rate of dehydrogenation of the secondary isopropyl alcohol, however, is five times as fast as that of the primary alcohol. This may be due to the fact that the secondary C—H linkage is normally weaker than the corresponding primary bond (cf. p. 588).

Catalytic Activity of Oriented Surfaces.—Interesting observations on the influence of the surface in catalysis have been made in connection with

the combination of hydrogen and ethylene on oriented films of metallic nickel. It has been found that by evaporating nickel in the presence of about 1 mm. pressure of an inert gas, e.g., nitrogen, the films deposited on glass are oriented so that the (110) planes of the crystals are largely parallel to the supporting material. By reducing the pressure of the inert gas the degree of order is diminished, and in a high vacuum the films obtained consist of randomly oriented crystals. From studies of the adsorption of various gases, it appears that the available surface of the oriented nickel film is twice that of the same weight of unoriented film, but the catalytic activity of the former for the hydrogenation of ethylene is ten times as great as that of the latter. It appears, therefore, that although the extent of adsorption on the (110) planes is not very different from that on the (100) and (111) planes of nickel, the catalytic activity of the (110) planes is considerably greater than that of the others. It is possible, in this particular instance, that the special properties of active centers may be associated with the distance of 3.51 Å. between nickel atoms which occurs mainly in the (110) planes and to a lesser extent in the (100) planes (O. Beeck, A. E. Smith and A. Wheeler, 1940).

Intermediate Compounds.—Several examples have been given of the formation of intermediate compounds as a factor in homogeneous catalysis, and there is evidence that they play a part in some heterogeneous reactions. The adsorbed layer of oxygen, for example, on a platinum catalyst, where forces of a chemical nature are probably involved, differs only from a definite intermediate compound in the respect that a separate phase is not formed, and the surface atoms of platinum still remain attached, to some extent, to the underlying atoms. If the surface atoms were detached, however, a definite oxide would probably result. This appears to happen when copper is used as a catalyst at 200° c. to bring about the union of hydrogen and oxygen. From a study of the rate of oxidation of copper to cuprous oxide by oxygen and the reduction of the oxide by hydrogen, strong evidence has been obtained that cuprous oxide is an intermediate in the combination of hydrogen and oxygen on copper; thus,



The direct combination of hydrogen and oxygen, necessitating a high energy of activation, is thus replaced by two processes each requiring a much lower activation energy. It should be emphasized that in this respect the reaction does not differ fundamentally from many other catalytic reactions in which there is no formation of an intermediate compound with a definite stoichiometric composition. The difference in behavior is one of degree rather than of type.²⁷

Other Functions of Surfaces.—Two other aspects of the influence of surfaces in gas reactions may be mentioned briefly: in the first place, a solid surface, e.g., the wall of the reaction vessel or other surface intro-

duced, may act as a means of removing energy in a three-body collision, and so facilitate the reaction (p. 1064). . Secondly, it has been seen that chain reactions are frequently initiated at the surface, although the actual propagation of the chain occurs in the gas phase. In a sense both these processes may be regarded as catalytic, since in the absence of the appropriate surface they would occur at a much slower rate.

Poisoning and Retardation.—In certain reactions the products, and even one of the reactants, are able to retard the process by being preferentially adsorbed and strongly held on the surface (pp. 1121, 1123). The rate is then inversely proportional to a simple power of the pressure of the retarding material. In other cases, purely extraneous substances, often present in very minute amounts, are able to inhibit catalytic reactions; such substances are known as **catalytic poisons**. It appears to have been known to many early investigators, e.g., E. Turner (1824), W. Henry (1824) and M. Faraday (1833), that hydrogen sulfide, carbon disulfide and other substances inhibited the activity of Döbereiner's catalyst, i.e., spongy platinum, for the combination of hydrogen and oxygen. The phenomenon is now known to be of frequent occurrence, and the poisoning effect of arsenic compounds on the platinum catalyst was the cause of the failure of the early attempts to establish the "contact process" for the manufacture of sulfur trioxide. It is of interest to note that the new vanadium oxide catalysts are not so susceptible to poisoning as is platinum. Poisons such as those mentioned above are permanent, in so far as they affect the catalyst to such an extent that it is virtually changed completely; chemical treatment for removal of the poison, however, will often regenerate the catalyst. A poison is generally very strongly adsorbed, and in extreme cases may even combine with the catalyst, so that the reactants are not able to reach the surface. Frequently very small amounts of poison, insufficient to form a complete unimolecular layer on the catalyst, have a considerable effect, as for example mercury and carbon monoxide in the combination of ethylene and hydrogen on a copper catalyst (cf. p. 1148). In these instances the poison appears to be preferentially adsorbed on the active centers, and so renders them inactive. Temporary poisons are effective only as long as they are present in the reacting gases. For example, when iron is used as a catalyst in the Haber process for the combination of nitrogen and hydrogen, water vapor and oxygen act as temporary poisons by uniting with the iron to form an oxide of iron. If the poison is removed from the gas the oxide is reduced to iron by the hydrogen present and its catalytic activity is restored.

Sometimes a catalyst can be poisoned for one reaction whereas another occurring on the same catalyst may be unaffected (cf. p. 1148); this fact can be turned to practical advantage. When ethyl alcohol vapor is passed over heated copper at about 300° c. the aldehyde first formed tends to decompose into methane and carbon monoxide. If water is present in the alcohol, however, the latter reaction is inhibited by partial poisoning of the catalyst, and the yield of acetaldehyde is thus increased.

Promoters.—The catalytic effect of a mixture is often greater than the sum of the separate effects of the constituents. In the extreme case a small quantity of a material, which is itself either noncatalytic or a feeble catalyst, is able to increase very appreciably the activity of a given catalyst. Such a substance is called a **promoter**; * many examples of promoter action in heterogeneous reactions are known and several are of industrial importance. In the combination of nitrogen and hydrogen the catalyst is iron, promoted by the addition of small amounts of potassium and aluminum oxides. The efficacy of iron for the catalytic oxidation of ammonia to form oxides of nitrogen can be increased by the addition of various metals, particularly bismuth; the actual catalyst is probably a mixture of oxides, and in fact the use of ferric oxide mixed with bismuth oxide has been patented (*B. A. S. F.*, 1914). This mixture is stated to form the basis of one of the catalysts employed in industry. The presence of nearly one per cent of ceria in the thoria used in making gas mantles is known to give the maximum efficiency, and it is of interest that one per cent of ceria gives the optimum promotion for the catalytic action of thoria on the combination of oxygen and hydrogen.

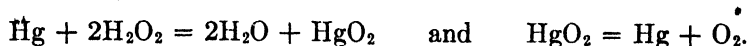
Mixed catalysts are often found to have enhanced activity; mention may be made of the substance "hopcalite," consisting of three parts of manganese dioxide to two of cupric oxide, which is a very effective catalyst for the union of carbon monoxide and oxygen at room temperatures, and of the mixture of zinc and chromium oxides, which is superior to either oxide separately for the reduction of carbon monoxide by hydrogen to yield methyl alcohol.

Attempts have been made to explain promoter action by the change in the spacing of the lattice and, presumably as a result, of the active centers of the catalytic surface, and also by changes in the adsorptive power of the catalyst, but none of these has proved entirely satisfactory. Heterogeneous reactions probably take place at discontinuities, and it may be that the promoter functions by increasing their extent. For example, on an ordinary iron catalyst one atom only in 2,000 appears to be able to catalyze the reaction between nitrogen and hydrogen, but when suitably promoted the proportion of active points is increased ten-fold. It must be remembered that just as there are many mechanisms of catalysis, so there may be various types of promoter action.

Heterogeneous Reactions in Solution.—The decomposition of hydrogen peroxide in aqueous solution in the presence of solid catalysts has been the subject of considerable study, particularly by G. Bredig, *et al.* (1899–1906). Colloidal metals, e.g., platinum and gold, are very effective catalysts, and the reaction is kinetically of the first order; this result, however, throws no light on the mechanism of the process. Massive metals are also able to catalyze the decomposition of hydrogen peroxide,

* Although instances of promoter action had been recorded previously, the word "promoter" was first used in the English specification of a patent taken out by the Badische Anilin and Soda Fabrik in 1910, in connection with the Haber process.

but as their effective area, per unit mass, is very much smaller, they are much less active than in the colloidal, or finely divided, state. When using mercury as catalyst it was observed that the decomposition was periodic and was accompanied by the regular formation and disappearance of a black film on the mercury surface; this black substance is apparently an unstable peroxide of mercury HgO_2 , and the catalytic action of the mercury may be due to the alternate formation and decomposition of this substance, viz.,



The view that higher unstable oxides act as intermediates in the decomposition of hydrogen peroxide is borne out to some extent by the fact that many metallic oxides are good catalysts for the process, the best being those capable of existing in two or more stages of oxidation, e.g., lead, manganese, silver, cobalt, nickel and gold oxides. The explanation advanced cannot, however, account for all the catalytic effects, for hydrogen peroxide is decomposed by dust particles, by glass wool and by many roughened surfaces, with which it does not form definite stoichiometric compounds. It is probable that adsorption of the peroxide results in the loosening of the bonds between the oxygen atoms, so that a relatively small energy of activation is sufficient to bring about decomposition.

The colloidal metals which act as hydrogen peroxide catalysts are particularly susceptible to the action of poisons, and in this respect there is a very marked similarity between the behavior of colloidal solutions and of enzymes capable of bringing about the decomposition of the peroxide, e.g., catalase. The results in Table 190, based on the work of

TABLE 190. POISONING OF COLLOIDAL PLATINUM AND BLOOD CATALASE

Poison	Volume containing 1 mole of Poison	
	Colloidal Platinum	Blood Catalase
Hydrogen sulfide	300,000 liters	1,000,000 liters
Hydrogen cyanide	20,000,000	1,000,000
Mercuric chloride	2,000,000	2,000,000
Mercuric cyanide	200,000	300,000
Iodine	5,000,000	50,000
Arsenous oxide	50	<2,000
Nitric acid	Non-poisonous	250,000

Bredig, give the dilution in liters at which 1 mole of the poison reduces the activity of colloidal platinum and of blood catalase to one half the initial value. With but few exceptions the parallelism is marked, and because of this Bredig applied the term "inorganic ferments" to colloidal solutions. The poisoning action on colloidal platinum is probably due to the preferential adsorption of the poison, so that the hydrogen peroxide does not have access to the surface. It is well known that certain substances, e.g., acetanilide, are sometimes added to hydrogen peroxide solution to stabilize it; this action may be due to preferential adsorption by, and

consequent poisoning of, dust particles and the glass walls of the containing vessel on which reaction chains are normally initiated. Concentrated solutions of the peroxide are kept in wax bottles, as this material is less catalytically active than glass. The stabilizers probably also act as chain breakers, and they may absorb light which produces photochemical decomposition of hydrogen peroxide.

The catalytic decomposition of hypochlorite solutions in the presence of the oxides of nickel or cobalt is a familiar reaction for the preparation of oxygen; the first stage in the process is apparently the formation of a higher oxide which decomposes to give oxygen, the original oxide being regenerated.

PHOTOCHEMISTRY

The subject of photochemistry is concerned mainly with chemical reactions resulting from exposure of a system to radiation, although various related topics are generally considered under the same head. The term radiation, strictly speaking, includes electromagnetic vibrations ranging from the electric waves of low frequency, through the infrared, visible and ultraviolet portions of the spectrum to the high frequency X-rays and γ -rays; the radiations of photochemical importance lie, however, almost exclusively in the visible and ultraviolet regions, i.e., from about 2000 Å. to 8000 Å. wave length. It has been known for many years that light is able to bring about chemical changes, but the systematic study of photochemistry was only commenced in the middle of last century. Progress was not rapid because of the lack of a unifying principle, but with the development of the quantum theory and the application of the results obtained from a study of molecular spectra, significant advances have been made. The subject of photochemistry is important both because of its intrinsic interest and because of the contributions it has made to the general problems of reaction kinetics.

Reactions of many types, e.g., synthesis, decomposition, hydrolysis, oxidation, reduction, polymerization and isomeric change, can be brought about by exposure to suitable light. Some photochemical reactions, however, differ in one respect from ordinary thermal, or "dark," reactions: whereas the latter are always accompanied by a decrease in free energy of the reacting system, certain photochemical processes involve an increase of free energy. Examples of such reactions are the ozonization of oxygen, the decomposition of ammonia, the polymerization of anthracene, and, one of the most important from the biological standpoint, the photosynthetic process occurring in plants whereby carbon dioxide and water, in the presence of chlorophyll, produce carbohydrates and oxygen. In all these cases some of the energy of the radiation is converted into free chemical energy of the products; when the source of light is removed, however, the system tends to return to its original state, although it may do so, as in the case of carbohydrates and oxygen, at a very slow rate at ordinary temperatures.

The Grotthuss and Draper Law.—The light absorbed and the chemical change occurring in a photochemical reaction are obviously closely related; the connection can be stated in terms of the generalization first made on theoretical grounds by T. von Grotthuss (1817), and later rediscovered by J. W. Draper (1841) as the result of experiments on the union of hydrogen and chlorine. It is generally known as the **Grotthuss-Draper law**, and may be stated as follows: *only radiations which are absorbed by the reacting system are effective in producing chemical change.* Although photochemical reaction can only result from the absorption of radiation, it does not follow that all the light absorbed is effective chemically; some of it may be converted into heat and some may be re-emitted as light of the same or another frequency. This emission of radiation is known by the general name of **fluorescence**.*

Fluorescence.—In order to understand the importance of the phenomenon of fluorescence to photochemistry it is convenient to consider an atom of an element or a molecule of a compound which does not decompose. When the gas or vapor is exposed to radiation of short wave length, i.e., high energy, an electronic transition will occur, and the atom or molecule is now in an excited state. If the additional energy is not removed by collision with another atom or molecule, then after the lapse of a short interval of time, about 10^{-7} or 10^{-8} sec., the electron returns to a lower level and radiation of frequency corresponding to the difference in energy of the initial and final electronic levels is emitted. It is this radiation which constitutes fluorescence. According to G. G. Stokes (1852) the radiation emitted in fluorescence has a greater wave length, i.e., lower energy, than the exciting radiation, that is the light originally absorbed. This result, which is found in many cases, is sometimes due to the fact that the absorbing molecules are nearly all in their lowest vibrational state, but after the emission of the fluorescence radiation they occupy higher vibrational levels. The energy of the emitted radiation will then be less than that of the absorbed light. A few instances are known of "anti-Stokes" behavior, when the fluorescence radiation has a higher energy than that absorbed. Since atoms do not possess vibrational energy, they exhibit what is known as **resonance fluorescence**, the fluorescent light having the same frequency as the exciting radiation. For example, if mercury vapor, with atoms in the normal 1S_0 state, is exposed to ultraviolet light, it will absorb resonance radiation of wave length 2537 Å., thus being converted into the excited 3P_1 state; if the pressure of the vapor is low, so that collisions are not too frequent, the electron in the excited atom will return to its normal level and emit the same radiation, i.e., 2537 Å.

Life of Excited Particles.—The life period of an excited atom or molecule can be estimated in various ways; the most direct is that of W. Wien (1924). Positive ions, produced in a discharge tube, are accelerated toward the cathode

* The name is derived from the fact that the mineral *fluorite* emits visible light on exposure to ultraviolet radiation.

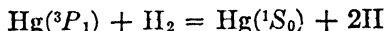
by a powerful electric field, and a proportion of them combine with electrons to form excited neutral atoms, or molecules, which pass through a small hole in the cathode at high speed. In returning to their ground state the particles give off the appropriate radiations in the visible or near ultraviolet parts of the spectrum, and hence a stream of luminous gas stretches from the back of the cathode into the attached observation chamber. The falling off in intensity of the luminosity of this gas stream depends on the rate of motion of the excited particles, and on their average life period. The connection between these quantities has been derived and since the intensity can be measured in a transverse direction, by means of a spectrophotometer, and the velocity of the particles is determined by the strength of the accelerating electric field, the life period of the excited atoms or molecules can be calculated. The values obtained vary with the nature of the return to the ground state. If the transition is a permitted one, as is commonly the case, the average life of the excited particle is usually about 10^{-7} to 10^{-8} sec. For a "forbidden" transition, however, the excited atom or molecule may have an average life of the order of 10^{-3} sec. or more. It is of interest to note that, in general, as a result of the operation of the uncertainty principle, the shorter the life period of an excited particle, the wider the spectral line due to the eventual transition to the ground state, for the shorter the life the greater the uncertainty in the energy.

Inelastic Collisions.—It has been assumed hitherto that the gas pressure is so low that the excited particle is not deprived of its energy by collision with another particle, either of the same or of a different chemical nature. At ordinary pressures one molecule collides with another once in approximately 10^{-10} sec., so that there should be many collisions during the life of an excited molecule; as a result, removal of energy will occur as the gas pressure becomes appreciable. This can be shown by the fact that inert gases are able to diminish or **quench** fluorescence. For example, hydrogen at a pressure of 0.2 mm. of mercury is able to reduce the resonance radiation of mercury vapor to one-half its original value, and although other gases are not so effective they still have a marked influence. As a general rule, the probability of energy transfer in a collision is greatest when the proportion converted into kinetic energy is least, the remainder being taken up in some other manner, e.g., chemical change or electronic excitation. When the excitation energy of the original atom or molecule is converted into kinetic energy, i.e., heat, the encounter is called a **collision of the second kind**, to distinguish it from one of the first kind, e.g., an encounter between an atom and a high velocity electron, in which kinetic is converted into excitation energy (see p. 51). From a study of collisions of the second kind it is possible to obtain an approximate estimate of the average life of an excited atom; if it can be assumed that every collision between the excited particle and a neutral one leads to deactivation, the life of the former can be calculated from the quenching effect of added gas at a known pressure. The results obtained in this way are very approximate, for deactivation does not always occur at every collision; this brings out the important fact that the transfer of electronic energy in collisions is highly specific, and depends both on the nature of the inert gas and of the excited particle. For example, hydrogen is more effective than nitrogen in quenching the resonance radiation of mercury vapor at 18°C ., but the reverse is the case for sodium vapor at 200°C . Nevertheless, by assuming that every collision with the most effective quenching gas is a deactivating one, an approximate, minimum value of the life period of the excited

particle can be obtained; this agrees with the results ascertained above. In some instances the deactivating influence of a substance is greater than would be expected from the number of collisions calculated from the kinetic theory employing the normal collision diameters, whereas in other cases it is less. The **quenching radius** or cross section of the deactivating molecule is thus generally not identical with the normal value.

Sensitized Fluorescence.—If the vapor of either thallium, zinc, cadmium, or an alkali metal is added to mercury vapor and exposed to the exciting radiation of 2537 Å. wave length, the resonance fluorescence radiation emitted contains lines of the spectrum of the added substance (G. Cario and J. Franck, 1922–23); the strongest lines are generally those of frequency nearest to that of the original light. The interpretation of this result, e.g., with thallium, is that the mercury vapor absorbs the 2537 Å. radiation and the atoms become excited; in collision with atoms of thallium, however, a proportion of the excitation energy is transferred to the latter, which are raised to higher electronic levels, viz., mainly 3S and 2D . The remainder of the energy is converted into kinetic energy. As the excited thallium atoms return to lower energy states the appropriate fluorescence radiation of thallium is emitted. This type of behavior is known as **sensitized fluorescence**; because of the presence of the mercury, the thallium is rendered fluorescent by radiations which would normally not be absorbed and would consequently have no effect. The electronic states of thallium which are nearest to that of mercury in the 3P_1 state are the 2S and 2D levels mentioned, and hence it appears, as indicated above, that the most probable energy transitions are those in which the maximum amount of electronic energy is transferred, leaving a minimum to be converted into kinetic energy.

Sensitized Decomposition.—Collision between a normal molecule and an excited particle may result in the transfer of so much energy to the former that it decomposes; an excellent illustration of this phenomenon is provided by the illumination of a mixture of mercury vapor and hydrogen gas by light of 2537 Å. wave length, when the hydrogen is found to be very active chemically. The properties of the gas show that it consists of atomic hydrogen, and that the process



has occurred. The essential probability of this dissociation is indicated by the fact that the difference of energy between the 3P_1 and 1S_0 states of mercury is 112 kcal. per g. atom,* whereas the heat of decomposition of a mole of hydrogen into normal atoms is about 103 kcal., leaving only 9 kcal. to be converted into kinetic energy. A process of this nature occurring as the result of energy transferred in a collision of the second kind is said to be a **photosensitized reaction**; further reference to this subject will be made later. It is of interest to record that cadmium

* For the calculation of the energy equivalent of radiation of known wave length, see p. 1160.

vapor gives a resonance radiation of wave length 3262 Å. which is effectively quenched by hydrogen; the molecules of the latter are not dissociated, since the quantum corresponding to the resonance line is only 88 kcal. The hydrogen molecule has, however, a vibrational level whose energy is close to that of the excited cadmium atom so that the transfer of energy occurs with very little converted into the kinetic form.²⁸

Fluorescence and Phosphorescence.—It is important to remember that both the exciting and emitted radiation may be in the ultraviolet region of the spectrum and hence not visible, as is the case with benzene vapor and other simple organic compounds. Many organic substances, however, undergo decomposition so soon after absorbing the exciting radiation that the extent of fluorescence is negligible. In the pure liquid state or in solution simple substances, inorganic and organic, are rarely fluorescent, presumably because there are so many deactivating collisions that the energy is almost entirely dissipated in the form of heat. A number of complex organic compounds of the fluorescein type, however, behave exceptionally in giving fluorescence in solution in the visible region of the spectrum. The extent of the fluorescence is influenced by the nature of the solvent, the effect of which varies from one substance to another, and by the presence of ions, particularly anions. Thiocyanate, iodide, bromide and chloride ions have a marked quenching influence; other ions act in the same way but are less effective. The proportion of the incident radiation emitted as fluorescence is increased by making the solution more viscous, for the number of deactivating collisions is reduced in this manner. The sodium salt of phenolphthalein, which does not show visible fluorescence in aqueous solution, fluoresces when dissolved in solid gelatin, sucrose, succinic acid or benzamide.

Phosphorescence* differs from fluorescence in the respect that the emission of radiation persists for some time, of the order of seconds or more, after the exciting radiation is cut off. Many dyes, which may exhibit fluorescence in the visible or ultraviolet regions in aqueous solution, phosphoresce when dissolved in a liquid, such as fused boric acid or glycerol, and cooled until it forms a rigid glass. In a medium of this type the number of deactivating collisions, which would remove the energy of the excited molecule, is very greatly decreased. A careful study of the phosphorescence of dyes shows that there are usually two distinct bands; for example, at ordinary temperatures fluorescein in boric acid exhibits phosphorescence in the blue (~5700 Å.) and yellow (~4800 Å.) regions. The former, known as α -phosphorescence, is identical with the fluorescence, although it persists for some time under normal conditions. As the temperature is lowered, however, the α -phosphorescence becomes less marked, and for fluorescein it disappears completely at 0° C. The yellow, or β -, phosphorescence is virtually independent of temperature, and it has been observed as low as -250° C. From measurements of the variation of the α -phosphorescence with temperature, the activation energy for the rate-determining stage in this process was found to be 8 ± 1 kcal., which is almost identical with the calculated energy difference for the blue and yellow phosphorescent radiations (G. N. Lewis, *et al.*, 1941).

The interpretation of these results is based on the suggestion of A. Ja-

* The name is derived from phosphorus which glows in the dark, although it is not strictly phosphorescent.

blonski (1935), that three energy states of the molecule must be considered, viz., the normal (N), the fluorescent (F), and the phosphorescent (P) states. The F state is merely the excited state of the molecule to which it is raised by absorption of energy from the N state, but the P state, which has a somewhat lower energy than the F state, is different in certain respects. It is possible that the phosphorescent state represents a stereoisomeric form, in which the configuration is held more or less fixed by the rigid solvent, but G. N. Lewis favors the view that it is a "triplet" or "bi-radical" state, that is, a state in which two electrons have uncoupled spins (cf. p. 619). The mechanism of phosphorescence is then as follows. Upon absorption of radiation the normal molecules become excited to the F state; they can then return, if not deactivated, to the N state, emitting fluorescence. In the rigid solvent, the extent of deactivation is decreased, and under certain conditions the molecule makes a transition from the F to the P state. The energy is small, e.g., 9 kcal. for fluorescein, and would correspond to radiation in the far infrared. If the P state is a triplet, as suggested above, the transition to the N state, which is a singlet, is "forbidden." The molecule in the P state will thus have a relatively long life. It is possible for it to return to the F state, but to do so it must acquire an activation energy equal to the energy difference in these states. Molecules which succeed in this respect pass back to the F state, and in returning to the N state emit the α -phosphorescence, which obviously will have the same frequency as the fluorescence. The necessity for an activation energy accounts for the fact that this type of phosphorescence is less marked as the temperature is lowered. Although the $P \rightarrow N$ transition is "forbidden," it does take place, and in this way the β -phosphorescence, of long life and independent of temperature, is obtained.²⁹

Law of the Photochemical Equivalent.—In the earlier photochemical studies attempts were made to correlate the extent of the chemical reaction with the amount of light energy absorbed in ergs or calories, but no simple relationship emerged from the results. An important advance was made by A. Einstein (1905) and J. Stark (1908), although its value was not realized until Einstein deduced the **law of the photochemical equivalent** thermodynamically (1912) and also on the basis of Bohr's model of the atom (1916). According to this law, sometimes called the Stark-Einstein law, *each molecule taking part in a chemical reaction induced by exposure to light absorbs one quantum of the radiation causing the reaction.* If the absorbing molecule decomposes or reacts immediately, without further successive or side reactions, then for every quantum absorbed one molecule should be involved in reaction.

If ν is the frequency * of the absorbed radiation, then the corresponding quantum is $h\nu$ (see p. 32); this is the amount of energy absorbed by each reacting molecule according to the Stark-Einstein law. The energy E absorbed by 1 mole is then $Nh\nu$, where N is the Avogadro number; thus,

$$E = Nh\nu = Nhc/\lambda,$$

* This is the true frequency in vibrations per sec.; to convert into $\bar{\nu}$ in wave numbers, divide by c , the velocity of light (see p. 37).

where c is the velocity of light and λ is the wave length of the absorbed radiation. Inserting the values for N as 6.023×10^{23} , h as 6.624×10^{-27} and c as 2.9977×10^{10} , the energy is obtained in ergs; dividing by 4.184×10^{10} gives the value in kcal., viz.,

$$E = 2.859 \times 10^5 / \lambda \text{ kcal. per mole,} \quad (191)$$

the wave length λ being expressed in Angström units. The quantity E for the given wave length is sometimes referred to as **one einstein** of radiation. It will be seen from (191) that the energy absorbed per mole is greater the shorter the wave length of the light; for violet light, for example, of wave length 4000 Å. the einstein is 71 kcal., whereas for red light of wave length 7500 Å. the corresponding energy absorption is approximately 38 kcal. From these results it is clear that radiations in the ultraviolet and violet portions of the spectrum should be more active chemically than those of longer wave length.

According to the law of the photochemical equivalent one mole of absorbing substance should decompose for every $2.859 \times 10^5 / \lambda$ kcal. of radiation absorbed, and the photochemical equivalent of 1 kcal. is thus $\lambda / 2.859 \times 10^5$ moles. This relationship between the amount of chemical action and the energy absorbed permits the law to be tested experimentally. The results are expressed in terms of the **quantum efficiency** or **quantum yield** ϕ , defined as the number of molecules decomposed by each quantum of radiation absorbed; thus,

$$\phi = \frac{\text{Number of molecules decomposed}}{\text{Number of quanta absorbed}}, \quad (192)$$

which should be equal to unity if the Stark-Einstein law is strictly obeyed. It was pointed out, however, by Stark, and more explicitly by M. Bodenstein (1913), that the law could be expected to apply only to the *primary* photochemical reaction, that is, the one in which the radiation was absorbed; the resulting activated, or even dissociated, molecule might be involved in secondary thermal reactions with the result that the quantum efficiency of the primary light-absorbing state might be obscured. If the secondary reactions are connected with the primary one in a simple stoichiometric manner, then the over-all quantum yield should be a simple integer, viz., 1, 2 or 3; results of this kind might, therefore, be regarded as confirming the fundamental reliability of the Stark-Einstein law and its strict applicability to the primary stage in which radiation is absorbed.³⁰

Experimental Methods.—For practical purposes the quantum yield may be expressed in the form

$$\phi = \frac{\text{Molecules decomposed per sec.}}{\text{Quanta absorbed per sec.}} = \frac{\text{Rate of chemical reaction}}{\text{Quanta absorbed per sec.}}, \quad (193)$$

the rate of chemical reaction being given in terms of molecules per second.

The experimental work falls, therefore, into two parts, viz., the determination of the rate of reaction, and the number of quanta absorbed, the latter involving the measurement of the energy and frequency, or wave length, of the absorbed radiation.

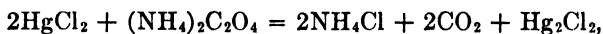
In the earlier work on photochemistry the reacting system was exposed to light without consideration of the wave lengths involved, but the development of the quantum theory has shown that results obtained in this way are of little value. It is now the practice, therefore, to use radiation, as far as possible, of a single definite wave length, i.e., monochromatic light, or covering a narrow range of wave lengths. The most common procedure is to employ a source of light, such as the quartz mercury-vapor lamp, giving a spectrum consisting almost entirely of a number of sharply defined lines; by means of suitable filters, consisting frequently of colored glass, liquid or gas, the unwanted radiations are filtered out, only the one required for the experiment emerging. Iron and carbon arcs are also used as light sources. For observations in the far ultraviolet the same principle can be adopted, except that a spark, e.g., between zinc or aluminum electrodes, is used as a source of illumination. For radiations in the visible or near ultraviolet not obtainable by means of the mercury-vapor lamp, a metal filament e.g., tungsten, lamp may be used in conjunction with a filter transmitting light over a narrow range of wave lengths only in the required region. For accurate work special monochromators have been used; they are somewhat similar to a spectrometer and consist virtually of a light source, such as described above, and a prism, of quartz if the radiation is in the ultraviolet region. In this way all light except that of the required wave length can be rejected.

The reaction cell is made of glass or quartz, depending on the radiation employed; the front and back should be plane parallel, and the parallel beam of incident radiation should impinge at right angles to these planes. The actual design of the cell will depend on the nature of the reaction, and whether it involves gas or liquid or both. The extent of the chemical change produced by light is relatively small; it has been calculated, for example, that radiation of 1 candle power falling on 1 sq. cm. corresponds to 2×10^{14} quanta per sec., so that this number of molecules should react per second. The number of moles reacting per sec. per candle power is therefore 3.3×10^{-10} per sq. cm. of substance exposed to radiation. It is obviously necessary to use as intense as possible a source of light and even then special methods must often be adopted to determine the rate of chemical change. If the reactants are gases and the reaction involves an increase or decrease of volume, the rate may be followed by measuring changes in pressure, e.g., by a mercury manometer, or with a Bourdon, Pirani or other gauge. It is sometimes possible, even when there is no volume change in the reaction, to determine the rate after condensing one or more of the components; for example, in the hydrogen-chlorine reaction, the chlorine and hydrogen chloride can be condensed by liquid air or other refrigerant, and the change in pressure of the residual hydrogen determined. Various other physical methods have been employed; examples are measurement of electrical conductivity, or of the change in the absorption of light due to the formation, or removal, of a colored substance, e.g., bromine. In other cases it has been necessary to resort to chemical analysis, but the methods must be capable of detecting small changes of concentration.

The amount of radiation absorbed is determined by measuring the intensity

before and after passing through the reacting system; this may be done by having two cells side by side, one empty, or containing the solvent if the reaction occurs in solution, and the other the reacting system. Alternatively, separate measurements may be made with the same cell empty and full; the difference gives the light absorbed by the reactants. In either of these ways errors due to light lost by reflection, or otherwise, are the same in both cases, and so cancel out. The actual intensity of the radiation, with and without the reacting system, is determined most accurately by a thermopile or a bolometer, generally the former; in each of these the radiation is converted into heat, and the amount is measured by the E.M.F. produced in the former, or by the change of resistance of a wire in the latter. The portion of the instrument exposed to radiation is blackened so that all wave lengths are absorbed. Calibration is carried out with the aid of a Hefner amyl acetate lamp or of a carbon-filament lamp previously standardized. Photoelectric cells are sometimes utilized for the measurement of light intensity; these have the disadvantage of different sensitivities for different wave lengths. At each particular wave length, however, the current produced by the cell is closely proportional to the intensity of the light falling on it, provided the accelerating voltage remains constant. Although the photoelectric cell is a convenient device it does not give a direct measure of the radiation intensity, for it must be calibrated for each wave length, against a thermopile or bolometer.

Frequently **actinometers**, in which the total *quantity* of light is determined by the amount of chemical reaction, are used in photochemical work. One of the earliest * of these, devised by J. W. Draper (1843), involves the hydrogen-chlorine reaction; it was improved by R. Bunsen and H. E. Roscoe (1857) and later by C. H. Burgess and D. L. Chapman (1906). The hydrogen chloride formed by exposure to light is absorbed by water and the contraction in volume is measured. In this, as in other actinometers, the assumption is made that the total extent of chemical reaction is proportional to the product of the light intensity and the time of exposure. In the actinometer of J. M. Eder (1879) use is made of the reaction



which occurs on exposure of the solution to light only if a trace of a ferric salt is present (see p. 1182); the amount of radiation falling on the solution is determined by the weight of mercurous chloride precipitated. In recent years the decomposition of oxalic acid in solution, sensitized by uranyl sulfate, has been studied with a view to its application in an actinometer for light of short wave length, e.g., in the ultraviolet; the amount of oxalic acid decomposed is determined by titration of the residual solution with permanganate. The chief disadvantage of chemical actinometers is that their efficiency, i.e., the extent of chemical reaction, varies with the wave length of the absorbed radiation; it is, therefore, necessary to standardize the reactions for different wave lengths. With the uranyl-oxalate actinometer, for example, the quantum yield at 25° c. decreases from 0.60 at 2540 Å. to 0.49 at 3660 Å., and then increases again to 0.58 at 4350 Å. (P. A. Leighton and G. S. Forbes, 1930).

From a knowledge of the wave length and the intensity of the absorbed radiation, that is, the amount absorbed per sec., the number of quanta absorbed

* The earliest actinometer is ascribed to T. de Saussure (1790); it involved the action of light on chlorine water and was used for his work on photosynthesis.

per sec. can be calculated, and if the rate of reaction is also known, in the appropriate units, the quantum efficiency of a process may be evaluated.³¹

Experimental Results.—Photochemical reactions fall into three main categories, according as the observed quantum yield is either (i) a small integer, (ii) very large, or (iii) very small. Reactions in the first group are those for which the law of the photochemical equivalent may be said to be applicable. It has been estimated that these constitute nearly 70 per cent of all reactions occurring under the influence of light; the law may thus be regarded as established experimentally. Some examples of reactions having quantum yields which are small integers are quoted in Table 191; the wave lengths of the radiations employed are given in each case.³²

TABLE 191. QUANTUM YIELDS OF REACTIONS OBEYING THE STARK-EINSTEIN LAW

Reaction	Wave Length	Yield
<i>Gaseous</i>		
$2\text{HI} = \text{H}_2 + \text{I}_2$	2070–2820 Å.	2
$2\text{HBr} = \text{H}_2 + \text{Br}_2$	2070–2530	2
$\text{SO}_2 + \text{Cl}_2 = \text{SO}_2\text{Cl}_2$	4200	~1
$\text{CH}_3\text{COCH}_3 = \text{C}_2\text{H}_6 + \text{CO}$	2500–3100	~1
$3\text{O}_2 = 2\text{O}_3$	1700–1900	3
$\text{H}_2\text{S} = \text{H}_2 + \text{S}$	2080	1
<i>Solution</i>		
$2\text{Fe}^{++} + \text{I}_2 = 2\text{Fe}^{+++} + 2\text{I}^-$	5790	1
$\text{CH}_3\text{C}(\text{I})\cdot\text{CO}_2\text{H} + \text{H}_2\text{O} = \text{CH}_3\text{OH}\cdot\text{CO}_2\text{H} + \text{HCl}$	2537	1
$\text{ClO}_2 = \frac{1}{2}\text{Cl}_2 + \text{O}_2$ (in CCl_4)	3000–4050	~2

It must be emphasized that there are many reactions for which the Stark-Einstein law does not appear to hold; for example, the quantum efficiency for the union of hydrogen and chlorine, in the absence of oxygen, is as high as 10^6 , and although values lower than 0.2 are not common for gas reactions, they are frequently found for photochemical processes occurring in solution. It is now generally agreed that the law of the photochemical equivalent is always applicable to the primary light-absorbing process, and that deviations are due to secondary processes. The high quantum efficiencies, for example, are ascribed to chain reactions, whereas the low values are accounted for by recombination of the products of the first stage of decomposition, and also by deactivation of the light-absorbing molecule by collision before it has time to react. The fact that very low quantum efficiencies are more common in solution, where the number of collisions is much greater, than in the gas phase supports this interpretation of the results. In the photochemical oxidation of sodium sulfite solution, in the presence of 0.1 M benzyl alcohol as inhibitor (p. 1142), 128 molecules of sulfite are oxidized for each quantum of radiation absorbed. At the same time one molecule of the alcohol is oxidized for 58 molecules of sulfite, so that if the former acts as chain terminator it may be supposed that the chain, following the primary absorption of light, involves 58 molecules of sulfite. It follows, therefore,

that $128/58 = 2.2$, i.e., approximately 2, reaction chains follow the absorption of each quantum. If two molecules of alcohol are oxidized when a chain is broken, the quantum efficiency of the primary process would be very close to unity, in agreement with the Stark-Einstein law. These results indicate clearly the necessity of distinguishing between the primary and subsequent stages of a photochemical reaction, and the two steps, viz., (a) the primary light-absorption process, and (b) the secondary chemical changes which follow, should be treated separately.

The Primary Light-Absorption Process.—Valuable information in connection with the primary process has been obtained from a study of molecular spectra; since exposure to infrared radiation, which only affects the rotational and vibrational energy of a molecule, rarely results in photochemical reaction, it is evident that such processes are almost invariably associated with electronic transitions, and consequently only electronic spectra need be considered. In Chapter VIII the analysis of molecular spectra was restricted to the type having bands with a definite fine structure; for photochemical purposes it is also important to examine spectra of molecules in regions where there is "continuous absorption," i.e., where no fine structure is evident, and also where the absorption is discontinuous but the band structure is diffuse.

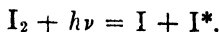
Continuous Absorption.—It is sometimes noted that the bands of discrete absorption become closer and closer together and are followed by a region in which no fine structure can be detected, and in which absorption is continuous; in the excited state, therefore, the vibrational and rotational levels no longer exist. The interpretation of this phenomenon given by J. Franck (1925) is that in the region of continuous absorption the molecule has dissociated into fragments, which may be atoms, radicals or molecules; the increase of vibrational energy accompanying the electronic transition is then so great that the vibrating parts break away from each other (cf. Fig. 128, p. 575). The energy corresponding to the convergence point of the discrete bands, that is, where the continuous absorption commences, may therefore be regarded as equal to the minimum energy required to bring about dissociation of the molecules in that particular electronic state. Any energy in excess of this value, corresponding to absorption at shorter wave lengths within the continuous region, will be generally converted into kinetic energy and will appear as heat. Absorption of radiation in the continuous portion of the spectrum is thus usually accompanied by photochemical dissociation.

The electronic spectrum of molecular iodine vapor consists of a series of fine-structure bands followed by continuous absorption; the convergence point of the former, at which the latter begins, is at 4991 Å. By means of (175) the corresponding energy per mole is given by

$$E = 2.859 \times 10^5 / 4991 = 57.3 \text{ kcal.},$$

assuming that each molecule of iodine absorbs one quantum of radiation. The thermochemical heat of dissociation of iodine is 35.5 kcal. per mole,

and the cause of the difference between this value and that obtained from spectroscopic measurements is that the excited iodine molecule does not decompose into two normal atoms, but into one normal and one electronically excited atom, viz.,



The arc spectrum of atomic iodine consists of doublets (cf. p. 62) due to the states $^2P_{3/2}$ and $^2P_{1/2}$, respectively, the energy difference of which is equivalent to 21.7 kcal. One of the atoms produced by the photochemical dissociation of iodine is in the normal $^2P_{3/2}$ state and the other in the excited $^2P_{1/2}$ level; the heat of dissociation into two normal atoms should be $57.3 - 21.7 = 35.6$ kcal. This result is in good agreement with the thermochemical value, and is evidence in favor of the view expressed that continuous absorption is accompanied by dissociation.

Calculations of the heat of dissociation of other molecules from the convergence limits of their spectra have been made in an analogous manner. It appears that dissociation of a covalent diatomic molecule results frequently in the formation of one normal and one excited atom, and consequently the appropriate allowance must be made in the calculations. Some of the results derived in this manner are recorded in Table 192; D^* refers to the heat of dissociation into one normal and one excited

TABLE 192. HEATS OF DISSOCIATION OF GASES FROM SPECTROSCOPIC DATA

Substance	Convergence Limit	D^* kcal.	D		
			Atomic Excitation kcal.	Spectro- scopic kcal.	Thermal kcal.
Chlorine	4785 Å.	59.7	2.5	57.2	57.0
Bromine	5107	55.9	10.4	45.5	46.2
Iodine	4991	57.3	21.7	35.6	35.5
Hydrogen	849	336.0	234.0	102.0	~100.0

atom, and D is that for two normal atoms. The convergence limit for hydrogen is seen to be in the far ultraviolet, so that only radiations of very short wave length will be able to cause the direct photochemical dissociation of molecular hydrogen.

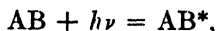
Although absorption of radiation in the continuous region brings about dissociation as the first step, it is important to remember that the ultimate yield of atoms may be small because of recombination. For example, since sunlight contains radiations of shorter wave length than the convergence limit for chlorine, the molecules of this gas should dissociate in daylight. Such dissociation undoubtedly occurs, but the resulting atoms unite so readily, in three-body collisions, that the equilibrium concentration of atoms is very small. If the apparent quantum yield were, therefore, to be evaluated by determining the atom concentration after the exposure of chlorine to light it would be found to be almost zero. It is this recombination of atoms which is responsible

for the **Budde effect**, the expansion observed on exposing chlorine or bromine vapor to light (E. Budde, 1873). The increase in volume is entirely due to the rise in temperature resulting from the heat evolved in the recombination of atoms. There is very little fluorescence, and almost the whole of the light energy is converted into heat by the consecutive dissociation and re-formation of the chlorine or bromine molecules.

Discontinuous Absorption with Diffuse Structure.—It was found by V. Henri (1924) that in certain electronic bands, e.g., sulfur (S_2) vapor, the discontinuous structure due to the quantization of vibration is present, but instead of the rotational fine structure the bands show a diffuse appearance. These are called **predissociation spectra**; they have been observed particularly for polyatomic molecules, e.g., sulfur dioxide, ammonia, nitrogen dioxide, acetaldehyde and methyl halides. The accepted interpretation of these diffuse spectra is that as a result of absorption of radiation in this region the molecule is raised to a higher electronic level, and at the same time there are vibrational transitions, so that the vibrational bands appear in the spectrum. With molecules having predissociation spectra it is supposed that within a short period the energy redistributes itself, leading to another excited state in which the available energy is sufficient to bring about dissociation. If the time elapsing between the acquisition of the necessary energy and its redistribution, resulting in dissociation, is small compared with the period of rotation of the molecule, about 10^{-11} sec., dissociation will take place before a complete rotation can occur and there will be no rotational fine structure (G. Herzberg, 1930). From the standpoint of photochemistry, the important conclusion is that the absorption of radiation in the predissociation region can lead to dissociation. However, the resulting atoms or radicals may recombine immediately, thus leading to a low over-all efficiency.

It may be noted that deactivation by collisions in the continuous and diffuse spectral regions is highly improbable, except at extremely high pressures. The time between acquisition of the excitation energy and dissociation is so small, viz., less than the period of a vibration or rotation, respectively, that the number of collisions under ordinary conditions is negligible.

Discontinuous Absorption with Fine Structure.—In the regions of discrete fine structure the radiation absorbed by a molecule serves to raise it from a lower to a higher electronic level, there being at the same time vibrational and rotational changes. The resulting molecule is then in an excited state, and the process of activation may be represented as



where $h\nu$ is the quantum of energy of frequency ν absorbed by each molecule of AB, and AB^* is the resulting activated or excited molecule. If the energy of the latter is insufficient to break a bond in the molecule AB, the

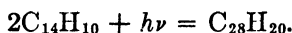
process is not of direct photochemical interest. The excited molecule will then lose its additional energy, within a period of 10^{-7} to 10^{-8} sec., either as the result of collisions with other molecules, which convert the electronic energy into translational energy, i.e., heat, or by the emission of the excess energy as fluorescence.

If the excited molecule possesses sufficient energy for dissociation to be feasible as a consequence of a rearrangement of the energy within the molecule, there is always a possibility that this will occur, either spontaneously or as the result of a collision with another molecule, which may be of the same kind or different. Since the life of an excited molecule is at least 10^{-8} sec., it may undergo many vibrations and rotations before dissociating, and so the spectrum will exhibit definite structure. It has been suggested that this type of behavior be referred to as "predissociation" (G. K. Rollefson and H. Burton, 1938), even though the absorbing molecule does not exhibit a predissociation type of spectrum. It is possible to distinguish between direct (or spontaneous) and induced predissociation, the latter requiring a collision with another molecule, by observing the effect of increasing the pressure or adding an inert gas. If these changes increase the photochemical yield, the rearrangement of energy within the molecule is induced; if the change were spontaneous, the yield would be unaffected. It must be remembered, however, that if an appreciable time is required for the rearrangement of the energy, there is an increasing probability that it will be lost either as fluorescence or in deactivating collisions. The former will tend to increase while the latter will decrease as the gas pressure is lowered.³³

Spectra of Liquids.—On account of the close proximity of the molecules in the liquid phase there are appreciable intermolecular effects which result in a broadening of the lines, so that they appear diffuse even in the portion of the spectrum in which discrete absorption occurs (p. 581). Formation of complexes and other types of interaction between solvent and solute lead to complications which do not arise in the spectra of gases. Nevertheless it is sometimes possible to distinguish between continuous and discontinuous absorption, and these correspond, as with gases, to dissociation and excitation, respectively, as the primary stage of the photochemical process. If the solvent does not interact in any way with the solute then the absorption spectrum of the solution shows the same general characteristics as that of the vapor, the energy relationships remaining almost unchanged. In spite of similarities in the absorption of radiation the quantum efficiencies may be different. In the liquid phase the probability of combination of two atoms as the result of a three-body collision is increased, and so also are the possibilities of deactivation of an excited molecule. The resulting quantum yields will thus, on the whole, be less than for the corresponding reactions in the gas phase. On the other hand, an excited molecule may have more opportunities for collisions leading to induced predissociation, thus increasing the efficiency of the process.

Secondary Chemical Reactions.—The nature of the reactions which follow the primary absorption of light determines the quantum yield, but the course of these reactions is governed entirely by ordinary thermal considerations. They are related to the primary process only because the latter provides the atoms, radicals or excited molecules taking part in the reaction; otherwise there is no connection between the two stages. Nevertheless, it is convenient to consider the secondary chemical reactions of various processes according to the three types of primary process described above.

Dimerization of Anthracene in Solution.—When anthracene, dissolved in benzene or other solvent, is exposed to ultraviolet light dimerization occurs and dianthracene is formed, viz.,



In dilute solution the reaction is accompanied by fluorescence and the quantum efficiency is small, but as the concentration is increased the

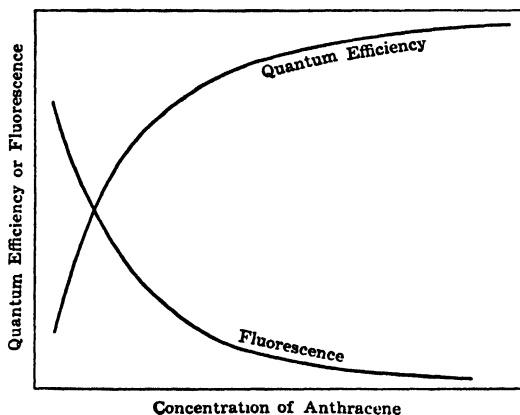
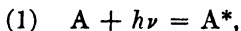


FIG. 253. Dimerization and fluorescence of anthracene in solution

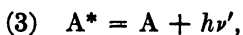
fluorescence falls off whereas the efficiency increases toward a limiting value; when the latter is attained there is very little fluorescence. The results are shown qualitatively by the curves in Fig. 253. This type of behavior is to be expected if the primary stage is activation of the molecule, i.e.,



followed by



where A represents anthracene and A₂ the dimeric form. When the concentration of anthracene is small, a large proportion of the excited molecules lose their energy as fluorescence of frequency ν' , viz.,



before they collide with other anthracene molecules. At higher concentrations, however, the number of collisions, leading to reaction and utilization of the excitation energy, is increased; the quantum yield consequently increases as the fluorescence radiation diminishes.

The activated molecules A^* are produced by process (1) and removed by (2) and (3). In view of their short life their concentration is small, and very soon after the reaction commences a stationary state is reached (p. 1079) in which the rate of formation of A^* by (1), which is proportional to the rate of absorption of radiation, is equal to the rate of removal by (2) and (3). Applying ordinary kinetic considerations, therefore, it follows that

$$k_1 I_{\text{abs.}} = k_2 c_A c_{A^*} + k_3 c_{A^*}, \quad (194)$$

where k_1 , k_2 and k_3 are constants, the two latter being the specific rates for processes (2) and (3), and $I_{\text{abs.}}$ is the rate of absorption of radiation, frequently called the intensity of the absorbed light.† If the value of c_{A^*} from (194) is inserted in the relationship

$$dc_{A_2}/dt = k_2 c_A c_{A^*} \quad (195)$$

derived from process (2), it follows that

$$\frac{dc_{A_2}}{dt} = \frac{k_1 I_{\text{abs.}} c_A}{c_A + k_3/k_2}, \quad (196)$$

which gives the rate of the photochemical reaction. In order to evaluate the quantum efficiency, it is necessary to divide by the rate of light absorption, i.e., by $I_{\text{abs.}}$ (cf. p. 1160); hence,

$$\phi = \frac{k_1 c_A}{c_A + k_3/k_2}. \quad (197)$$

From this equation it is clear that ϕ increases as c_A increases, as found experimentally; a limiting value of the quantum yield is reached at high concentrations when k_3/k_2 can be neglected in comparison with c_A . It may be seen from reactions (1) and (2) that the maximum number of molecules of A_2 formed for each quantum of radiation absorbed is one, so that the maximum quantum yield should be unity. The experimental value appears to be smaller than this, so that the mechanism described does not give a complete representation of the secondary process; deactivating collisions of the "second kind" with solvent molecules may be a factor to be considered. Mention should also be made of the fact, which has been ignored in the foregoing treatment, that thermal dissociation of A_2 takes place (see p. 1184). This will tend to decrease the observed quantum yield, but the effect will be small provided the concentration of A_2 , relative to A , remains small.

† According to Beer's law (p. 581), $I = I_0 e^{-\alpha c d}$, so that $I_{\text{abs.}} = I_0 - I = I_0(1 - e^{-\alpha c d})$. If the extent of absorption is small, α is small, and $I_{\text{abs.}} \approx I_0 \alpha c d$; on the other hand, if absorption is considerable, and α is large, $I_{\text{abs.}}$ is almost identical with I_0 , the intensity of the incident radiation.

Decomposition of Hydrogen Iodide.—The photolysis† of hydrogen iodide is an example of a photochemical process brought about by radiation in the region of continuous absorption, and so involves dissociation into atoms as the primary stage. Incidentally, the reaction has provided one of the most striking illustrations of the applicability of the Stark-Einstein law. E. Warburg (1918) studied the decomposition in ultra-violet light of wave lengths 2070, 2530 and 2820 Å.; the results obtained are given in Table 193, where the second column gives the number of

TABLE 193. PHOTOCHEMICAL DECOMPOSITION OF HYDROGEN IODIDE

Wave Length	Moles per Kcal.	Yield
2070 Å.	1.44×10^{-2}	1.98
2530	1.85	2.08
2820	2.09	2.10

moles of hydrogen iodide decomposed per kcal. of radiation absorbed, and the third the quantum yield, that is, the number of molecules decomposed per quantum. It is evident that whereas there is no simple connection between the extent of photodecomposition and the amount of energy absorbed, the case is quite different if the photolysis is calculated per quantum of radiation; the yield is two, within the limits of experimental error, for the three wave lengths studied. The quantum yield is unaffected by decreasing the pressure of hydrogen iodide to 0.1 mm. of mercury, or by the presence of inert gas, viz., nitrogen; in liquid hydrogen iodide the yield is only slightly reduced, to 1.84.

The spectrum of hydrogen iodide is continuous over a considerable range of wave lengths, and in the region for which the results in Table 193 were obtained the primary stage is dissociation; thus,



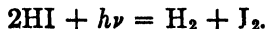
a normal hydrogen and an excited iodine atom being formed. The products may then take part in the following secondary reactions:



and



Reactions (4) and (6) are so highly exothermic, viz., 100 kcal. and 68 kcal., respectively, that they will only occur in three-body collisions, and reaction (3) is endothermic to the extent of at least 11 kcal., and possibly more if the excited iodine atom has lost some energy in collisions, so that it will require a relatively high activation energy. It appears, therefore, that reactions (2) and (5) are the most important secondary chemical processes, and that the net result is



† The term photolysis is frequently used for photochemical decomposition.

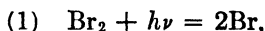
The yield should thus be two molecules of hydrogen iodide decomposed per quantum, as actually found. The excited iodine atom produced in the first stage is of no special importance, as similar results have been obtained with radiation, viz., 4047 Å. wave length, which gives a normal iodine atom. The reaction



like the corresponding ones with chlorine and bromine (pp. 1081, 1079), can take place at an appreciable rate, the activation energy being small. However, because the concentration of molecular iodine is low, this reaction has little effect on the efficiency of photolysis of hydrogen iodide at the beginning of the decomposition. In the later stages it tends to decrease the effective yield.

The quantum yield for the photolysis of hydrogen bromide is also two, and spectroscopic measurements indicate that the mechanism of the process is exactly analogous to that for hydrogen iodide.

The Hydrogen-Bromine Reaction.—There is a marked similarity between the photochemical combination of hydrogen and bromine and the thermal reaction (p. 1079), and so it is probable that the secondary chemical stages obey the ordinary laws of chemical kinetics. When the photoreaction is brought about by light in the continuous region of the bromine spectrum, the first stage is



so that the rate of formation of bromine atoms in this step is equal to $k_1 I_{\text{abs.}}$, where k_1 is a constant and $I_{\text{abs.}}$ is the rate of absorption, i.e., intensity, of radiation. The subsequent chemical changes are then identical with (2) to (5) of the thermal reactions on p. 1079, and the rate of formation of hydrogen bromide is given by (82). In fact the two processes may be regarded as identical, but for the fact that in the first stage of the thermal reaction the rate of formation of bromine atoms is $k_1 c_{\text{Br}_2}$, compared with $k_1 I_{\text{abs.}}$ for the photochemical process. Accepting the view that a stationary state is soon reached in the photoreaction, it is a relatively simple matter to derive the equation

$$\frac{dc_{\text{HBr}}}{dt} = \frac{2k_2 \sqrt{k_1 I_{\text{abs.}}/k_5 c_{\text{H}_2}}}{1 + \frac{k_4 \cdot c_{\text{HBr}}}{k_3 \cdot c_{\text{Br}_2}}} = \frac{k \sqrt{I_{\text{abs.}} c_{\text{H}_2} c_{\text{Br}_2}}}{k' c_{\text{Br}_2} + c_{\text{HBr}}}, \quad (198)$$

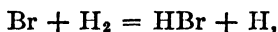
which agrees with the experimental data of M. Bodenstein (1924). It accounts for the surprising fact that the reaction rate is proportional to the square root of the light intensity. The mechanism postulated applies strictly at constant pressure, and if this is varied the denominator of (198) should be multiplied by the square root of the total pressure. This result is readily interpreted by assuming, as is reasonable, that the chain-terminating reaction (5), on p. 1079, i.e., the combination of bro-

mine atoms, involves a three-body collision, viz.,



where X is any other molecule. If (5) is replaced by (5a) the quantity $\sqrt{c_X}$ is included in the denominator of (82), and since X may be any molecule this term is equivalent to the square root of the total pressure. The new relationship only holds, however, if the pressure is not too low, for otherwise the rate of formation of hydrogen bromide would apparently approach infinity as the total pressure was diminished. At low pressures the union of bromine atoms occurs mainly at the walls of the vessel, and so reaction (5a) is not operative. It is of interest that at low pressures inert gases, e.g., carbon dioxide, nitrogen, oxygen, argon, etc., are able to increase the reaction velocity; they prevent the bromine atoms from reaching the walls where they would combine and thus terminate the chains.

In spite of the chain mechanism the quantum yield of the photochemical hydrogen-bromine combination is very small, viz., about 0.01, at ordinary temperatures, although it increases as the temperature is raised. The reason for this is that the reaction immediately following the primary stage, i.e.,



is endothermic and so requires a high energy of activation (p. 1080). At ordinary temperatures this reaction is so slow that most of the bromine atoms recombine to produce molecules, thus giving a low yield. As the temperature is raised, however, the rate of reaction is increased and so also is the quantum efficiency of the whole process.

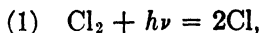
Another interesting feature of the photosynthesis of hydrogen bromide is that the rate of reaction is the same in the discontinuous as in the continuous portion of the spectrum. This means that bromine atoms can be formed by direct dissociation of the molecules in the continuous region or by predissociation, probably induced, in the banded region, viz.,



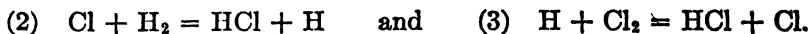
where X is any other molecule. If every excited Br^* molecule decomposes, the rate of formation of bromine atoms will be proportional to the light absorbed, just as when direct dissociation occurs. Other instances of this type of behavior, namely the same quantum yield and reaction velocity, on both sides of the convergence limit are known; for example, two molecules of ozone are formed per quantum of energy absorbed by oxygen gas at wave lengths of 2070 Å. and 1700 Å., although the continuum commences at 1750 Å. It must be emphasized that the range of wave lengths over which the yield remains unchanged is not very large; the molecule must absorb enough energy to dissociate immediately, i.e., in the continuum, or it must be so close to this amount that dissociation occurs readily either spontaneously or in a collision.

The Hydrogen-Chlorine Reaction.—One of the most interesting, if perplexing, photochemical processes is the combination of hydrogen and chlorine; it was first observed by W. Cruickshank (1801), and confirmed by others soon afterward. The earliest detailed studies of the reaction were made by J. W. Draper (1841–45) and particularly by R. Bunsen and H. E. Roscoe (1855–59), and from this work there emerged the following facts, which were difficult to understand at the time but can now be readily explained. First, the combination of hydrogen and chlorine was preceded by a period of induction (p. 1077), which has been shown to be due to the presence of nitrogenous impurities such as ammonia (C. H. Burgess and D. L. Chapman, 1906). These substances act as inhibitors by breaking the reaction chains, and only when they are completely converted into nitrogen and ammonium chloride can the normal photosynthesis of hydrogen chloride take place. With pure hydrogen and chlorine no induction period is observed. It may be mentioned that oxygen also acts as an inhibitor, but its action is different from that of the nitrogen compounds; it does not prevent the reaction from starting, but slows down its rate. The second discovery was that the commencement of the hydrogen-chlorine combination is accompanied by an increase of volume at constant pressure; this has been called the **Draper effect**. Since two molecules of hydrogen and chlorine give two molecules of hydrogen chloride the volume should remain constant, and the increase has been traced to the heat evolved in the chemical reaction which causes a rise of temperature and consequently an expansion. It has been frequently stated that thoroughly dry hydrogen and chlorine do not unite photochemically; this claim has now been disproved, so that it need not be considered further (cf. p. 1129). It is probable that impurities which inhibit the reaction were introduced in the process of drying.

The quantum yield of the hydrogen-chlorine reaction is exceptionally high; it varies somewhat with the conditions, but values from 10^4 to 10^6 are not uncommon in the absence of oxygen, with light of wave length 4800 Å. or less.* The suggestion was made, first in a somewhat vague manner (1913) and later more explicitly (1916), by M. Bodenstein that this result could be explained by a chain reaction, and the basis of the mechanism now universally accepted was proposed by W. Nernst (1918). When exposed to light in the continuous region of the chlorine spectrum, i.e., wave length 4785 Å., the first stage in the photochemical reaction is the dissociation of chlorine; thus,



and then follow the reactions



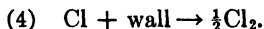
* The process occurs in the range 5460 Å. to 4800 Å. with an apparently lower efficiency, because only a small proportion of the light absorbed is capable of causing the chlorine molecules to dissociate, and this is the first stage in the reaction.

The regeneration of the chlorine atom, after the production of two molecules of hydrogen chloride, permits reactions (2) and (3) to occur again, and so the chain is propagated. At first sight it appears that the mechanism is the same as that for the hydrogen-bromine reaction, for which the efficiency is very low; the difference lies in the fact that reaction (2) between chlorine atoms and hydrogen molecules is very rapid, since it is exothermic and has a relatively low activation energy, whereas the corresponding reaction with bromine atoms is very slow at ordinary temperatures (p. 1172).

There is much evidence in favor of the chain mechanism apart from the high quantum yield; a portion only need be given here. The combination of hydrogen and chlorine can be initiated, without exposure to radiation, by introducing either chlorine or hydrogen atoms into the gas, and the results are similar to the photochemical reaction. Further, the quantum yield of the latter is markedly decreased by working in capillary tubes, since the chains are terminated at the walls; a similar reduction is observed at low pressures, for the same reason.

Although there is general agreement on the essentials of the chain mechanism, there is still, in spite of many investigations, uncertainty as to the processes whereby the chains are terminated; these are important since they determine the kinetics of the reaction, and the dependence of the velocity and the quantum yield on the light absorbed. One point, however, is clear: the reaction mechanisms in the presence and absence of oxygen are essentially different, and consequently they must be considered separately.

The chief difficulty in interpreting the results of the hydrogen-chlorine reaction is their contradictory nature; this is due in a great measure to faulty experimental technique leading to erroneous conclusions. In spite of the confusion, there is general agreement on certain points and one of the most important is that, in the absence of oxygen, provided the chlorine pressure is not too low, the rate of reaction is proportional to the intensity of the absorbed light and to the pressure of hydrogen. This result follows directly from the chain mechanism given above with the additional assumption that the chains are broken by the combination of chlorine atoms, to form molecules, on the walls of the reaction vessel; thus,



Chlorine atoms are formed in reactions (1) and (3), given on p. 1173, and are removed by (2) and (4), so that in the stationary state

$$k_1 I_{\text{abs.}} + k_3 \text{CHCl}_2 = k_2 \text{CClCH}_2 + k_4 \text{CCl}, \quad (199)$$

the assumption being made that (4) is a first-order reaction. Similarly, considering the stationary state for hydrogen atoms, it is seen that

$$k_2 \text{CClCH}_2 = k_3 \text{CHCCl}_2, \quad (200)$$

and hence from (199) and (200),

$$k_1 I_{\text{abs.}} = k_4 c_{\text{Cl}}; \quad (201)$$

$$\therefore c_{\text{Cl}} = k_1 I_{\text{abs.}} / k_4. \quad (202)$$

The total rate of formation of hydrogen chloride, by reactions (2) and (3), is represented by

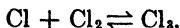
$$\frac{dc_{\text{HCl}}}{dt} = k_2 c_{\text{Cl}} c_{\text{H}_2} + k_3 c_{\text{H}} c_{\text{Cl}_2} \quad (203)$$

$$= 2k_2 c_{\text{Cl}} c_{\text{H}_2}. \quad (204)$$

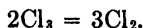
Introducing the value for c_{Cl} given by (202), it follows that

$$\frac{dc_{\text{HCl}}}{dt} = \frac{2k_1 k_2}{k_4} I_{\text{abs.}} c_{\text{H}_2} \quad (205)$$

in agreement with experiment. At high chlorine pressures the chain-terminating reaction (4) may possibly be replaced by the equilibrium

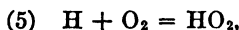


followed by

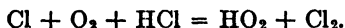


This, of course, leads to a kinetic expression different from (205).

For the hydrogen-chlorine reaction in the presence of a small amount of oxygen the simplest assumption to make is that the normal stages (1), (2) and (3) are followed by the main chain-terminating process



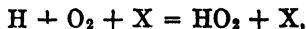
the radical HO_2 subsequently reacting to give products which do not propagate chains. Some chains are apparently also broken by process (4), given above, and by the reaction



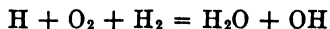
It is seen that while in the absence of oxygen it is necessary to suppose that chains are destroyed by the removal of chlorine atoms, in the presence of oxygen it is postulated that the removal of hydrogen atoms largely determines the kinetics of the reaction. From reactions (1), (2), (3) and (5) it may be deduced, using the concept of the stationary state for the concentration of hydrogen and chlorine atoms, that

$$\frac{dc_{\text{HCl}}}{dt} = \frac{kc_{\text{Cl}_2} I_{\text{abs.}}}{c_{\text{O}_2}}, \quad (206)$$

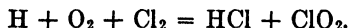
assuming (5) to be the only chain-ending process. The rate of reaction in the presence of oxygen is in fact directly proportional to the light intensity, and the inhibiting action of the gas is approximately proportional to its concentration, as required by (206). The actual kinetics of the process are, however, very much more complicated than this equation would imply, and it appears that process (5) represents an over-simplification. In the first place it probably requires a three-body collision, viz.,



where X is any molecule capable of removing some of the energy of the reaction, and in the second place, X may be a molecule, e.g., H_2 or Cl_2 , capable of reacting, viz.,

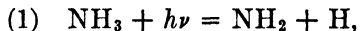


and

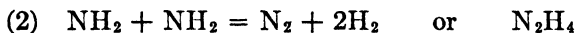


Measurements of the amounts of water and hydrogen chloride formed during the reaction suggest that both these processes occur. Incidentally, it may be noted that the production of water when a mixture of hydrogen, chlorine and oxygen is exposed to light is a photosensitized reaction (p. 1180), for the radiations are actually absorbed by the chlorine molecules.³⁴

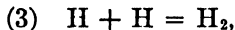
Photolysis of Ammonia.—Ammonia gas gives a diffuse, predissociation type of spectrum in the region 1600 to 2200 Å., and such radiation is effective in decomposition; this should presumably involve dissociation in the primary stage, probably



for the presence of hydrogen atoms can be proved by the influence of the photochemical system on the conversion of para- to ortho-hydrogen (p. 1060). The quantum yield varies with the temperature, e.g., 0.15 at 20° c. to 0.5 at 400° c. with light of 2144 Å., and increases and then decreases with increasing pressure. The low quantum yield is not due to deactivating collisions, because it remains small at very low pressures, and so it must be attributed to recombination of the products of process (1). It is significant that the introduction of hydrogen atoms from an external source retards decomposition of the ammonia still further. The final products of photolysis are nitrogen, hydrogen and hydrazine, resulting from such stages as



and

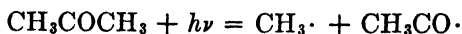


and it is the relative extent to which these reactions, and the reverse of (1), take place under different conditions that determines the influence of temperature and pressure on the quantum yield. It is probable, also, that all three processes require three-body collisions, thus introducing a further variable factor.

Aldehydes and Ketones.—The photolysis of aldehydes and ketones has attracted much attention in recent years. These substances show discontinuous absorption, at wave lengths greater than about 3000 Å., which then passes into a diffuse, i.e., predissociation, region, or through a short diffuse portion into a continuum. This region of continuous absorption is characteristic of the carbonyl group, but it is nevertheless not the $C=O$ bond which is broken. In the first place no oxygen has been detected in the photolysis of aldehydes and ketones, and secondly, the energy

corresponding to radiation of 3000 Å. is about 95 kcal. per mole [equation (191)] which is insufficient to disrupt the double linkage (p. 590). Dissociation does, however, occur as is shown by the ready formation of free radicals on exposure of the vapors of acetaldehyde, acetone, and methyl ethyl, dipropyl and other ketones to radiation of short wave length. It appears, therefore, that although the radiation is absorbed by the carbonyl group, it passes in some manner to an adjoining C—C linkage; this requires about 60 kcal. for dissociation and so the available energy is sufficient to cause it to break.

The decomposition of acetone vapor by radiation in the continuum yields mainly carbon monoxide and ethane, together with some methane, as the gaseous products, and also appreciable amounts of diacetyl. These results, together with the fact that the presence of free radicals has been proved, suggest the primary stage *



Numerous secondary reactions are possible, but a few only need be mentioned here. The diacetyl is formed by the combination in pairs of acetyl radicals, i.e., $\text{CH}_3\text{CO}\cdot$, and carbon monoxide and ethane are produced by the reactions



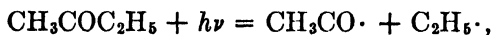
The formation of ethane from two methyl radicals, requiring a three-body collision, is apparently not very probable. The small amounts of methane which have been detected are presumably formed by the interaction of methyl radicals with acetone molecules. It will be observed that there is no chain mechanism, and so the quantum yield is of the order of unity.

Acetone vapor also undergoes decomposition by light of wave length over 3000 Å., e.g., 3130 Å., which is in the region of the spectrum with discrete structure. The energy is clearly sufficient to bring about dissociation of a C—C bond, and this presumably occurs some time after excitation; the initial products appear to be free radicals, and the subsequent products, viz., carbon monoxide, ethane and diacetyl, are as described above. The essential difference between the photochemical processes in the continuum and in the discrete region is that in the latter case an appreciable time, greater than 10^{-13} sec., but less than about 10^{-8} sec., elapses between the absorption of radiation and actual dissociation. It is not surprising, therefore, that the quantum yield is less in the discrete spectral region, because of the greater probability of loss of energy by fluorescence and deactivating collisions.

The photolysis of methyl ethyl ketone in the continuous portion of the spectrum must be similar to that of acetone, and the presence of

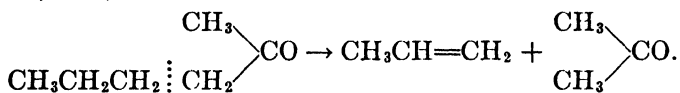
* Strictly speaking, the primary stage is excitation of the molecule, in the C=O bond, followed by rearrangement of the energy leading to dissociation within less than the time of a complete vibration.

free radicals has been established. From the nature of the products it appears that the first stage is probably



the essential mechanism, i.e., absorption by the C=O group and rearrangement of the energy, being the same as for acetone. The $\text{CH}_3\text{CO}\cdot$ radical then decomposes to yield methyl radicals and carbon monoxide. Reaction of $\text{CH}_3\cdot$ and $\text{C}_2\text{H}_5\cdot$ with $\text{CH}_3\text{CO}\cdot$ accounts for the formation of ethane and propane, respectively. Some butane may result from combination of ethyl radicals in pairs.

In addition to the type of decomposition already considered, ketones containing a relatively long chain, e.g., methyl *n*-butyl ketone, or certain cyclic ketones, undergo an interesting photochemical change. The energy absorbed by the C=O group is not transferred merely to an adjacent linkage, but to one beyond, namely that joining the α - and β -carbon atoms; thus,



It is possible that no free propyl radicals are formed in the primary photochemical stage, but there is a valence rearrangement in the molecule so that propylene splits off as such.

The decomposition of aldehydes under the influence of ultraviolet light has, as may be expected, many features in common with the photolysis of ketones; there is, however, one important difference. The quantum yield for the latter does not become appreciably greater than unity, but with aldehydes high yields are obtained at elevated temperatures, indicating that a chain mechanism becomes operative (J. A. Leermakers, 1934). The results in Table 194 (R. G. W. Norrish, 1936) were obtained

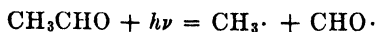
TABLE 194. QUANTUM YIELDS IN LIGHT OF 2500-3100 Å. WAVE LENGTH

Temperature	Formaldehyde	Acetaldehyde	Acetone
100° c.	1.0	1.0	0.47
200°	2.9	11.5	0.5
300°	35.0	70.4	0.65
400°	100.0	138.0	0.7

under comparable conditions for gaseous formaldehyde, acetaldehyde and acetone with light between 2500 and 3100 Å. The difference in behavior between aldehydes and ketones is connected with the reactive hydrogen of the $-\text{CHO}$ group of the former.

Various suggestions have been made concerning the nature of the chain process occurring in aldehydes upon photolysis. The kinetics of the reaction with acetaldehyde at high temperatures can be accounted for by postulating, as the first stage, absorption of radiation followed by rearrangement within

less than the time of a vibration, leading to dissociation, viz.,

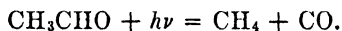


This is seen to be analogous to stage (1) in the proposed mechanism for the thermal reaction, given on page 1084; if the subsequent stages are taken to be the same as described there, the chain-terminating reactions (7) and (8) being neglected, as before, it is readily found that the rate of photolysis of acetaldehyde, as determined by the rate of formation of carbon monoxide, is given by

$$\frac{dc_{\text{CO}}}{dt} = kI_{\text{abs.}} + k_1c_{\text{CH}_3\text{CHO}}I_{\text{abs.}}^{\frac{1}{2}} \quad (207)$$

where k and k_1 are constants. The observed kinetics at higher temperatures are in agreement with (207). If the rate of formation of carbon monoxide is divided by the intensity of the absorbed radiation, the result is the quantum yield of the process; this is seen to vary with the light intensity, as is frequently the case with photochemical chain reactions.

The chain carrying process (5), page 1084, has an appreciable energy of activation, e.g., about 10 kcal. This may account for the fact that the quantum yield decreases as the temperature is lowered. There appears to be a change in the reaction mechanism at the same time. At low temperatures, when the efficiency is unity, it is possible that the $\text{CH}_3\cdot$ radicals and H atoms formed in the first stages of decomposition combine immediately, before they leave their mutual spheres of influence, so that the net reaction is



This type of decomposition is apparently favored by the longer wave length radiations; the smaller the energy of the excited molecule, the smaller will be the probability that $\text{CH}_3\cdot$ and H will have sufficient kinetic energy to escape from each other.

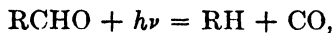
It is important to note that both formaldehyde and acetaldehyde, like acetone, can be decomposed by light in the discontinuous, fine-structure region of the spectrum provided the wave length is not very much greater than that required to bring about direct dissociation. The mechanism of the process is essentially the same as that described above for acetone. The quantum yield, at low temperatures, is less than unity, because of deactivating collisions and emission of radiation as fluorescence.⁴²

Liquid Phase Reactions.—On the whole, photochemical reactions in the liquid phase have not been investigated as thoroughly as those in the gas phase, partly because of the associated complexities; there is, as a general rule, no fundamental difference between the processes in the two phases. The quantum yield for the decomposition of hydrogen iodide is almost the same for the gas as for the pure liquid, and for a molar solution in hexane, although there is some falling off in the more dilute solutions and at the shorter wave lengths, indicating that the process is not altogether simple in the liquid phase. The photolysis, as in the gas reaction, occurs in the region of continuous absorption, and involves dissociation into atoms as the primary stage. The yields for the photolysis of chlorine monoxide, of nitrogen trichloride, and of ozone sensitized by

chlorine (p. 1182), are almost the same in carbon tetrachloride solution as in the gaseous phase, but photochemical reactions taking place in the liquid phase are not always as simple as those involving gases.

The mutual interconversion of fumaric and maleic acids occurs in aqueous solution when exposed to ultraviolet light, viz., 2070 to 2820 Å. The action of the light is not merely to facilitate the attainment of a normal equilibrium, for fumaric acid is the stable form at ordinary temperatures, but on exposure to light the system tends to 75 per cent maleic acid, the actual proportion varying with the concentration and other conditions. It has been suggested that absorption of energy by the fumaric (*trans*) or maleic (*cis*) acid molecule raises it to an excited state in which the restriction to internal rotation about the double bond is less than in the normal state. If the molecule then acquires additional energy of internal rotation, e.g., in a collision or in other ways, the change from *cis*-form to *trans*-form, or *vice versa*, can take place. The possibilities of return of the molecule to the ground state, before complete rotation occurs, and of deactivating collisions are relatively great, and hence the quantum yields are small, e.g., 0.03 to 0.05 for the conversion of maleic to fumaric acid, and 0.09 to 0.13 for the reverse change. The photochemical *cis-trans* isomeric changes of cinnamic acid, stilbene and azobenzene in solution, are presumably similar to that described above.

The photolysis of aldehydes and ketones in solution is, in the main, similar to the decomposition in the gas phase. The two chief processes, viz., (a) formation of carbon monoxide and paraffin hydrocarbons, and (b) the production of an olefin and a simpler aldehyde or ketone, have been observed. Some complication is introduced, however, by interaction of the free radicals formed in the primary stage of the photolysis of ketones with the hydrocarbon solvent. In the photochemical decomposition of aldehydes at temperatures up to 100° c. no such reaction is observed; it appears, therefore, that decomposition occurs directly into molecules, viz.,

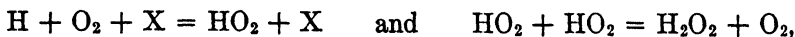


without the intermediate production of free radicals, or if such radicals are formed they must interact immediately (cf. p. 1179). Photolysis in solution takes place less readily than in the gas phase; acetone does not decompose appreciably up to 80° c. in a hydrocarbon solvent, and with other ketones a temperature of at least 60° c. is necessary for the decomposition to be studied (R. G. W. Norrish, 1938).³⁶

Photosensitization.—The earliest studies of photosensitization in gas reactions were made by F. Weigert (1907) who observed that in the presence of chlorine photochemical reactions between sulfur dioxide and oxygen, and between hydrogen and oxygen (p. 1175), and decomposition of phosgene and ozone, occurred under the influence of blue or violet light; in the absence of the light-absorbing sensitizer, chlorine, little or no reaction took place. It has been already recorded (p. 1157) that in the pres-

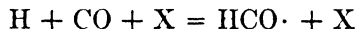
ence of mercury vapor hydrogen can be dissociated by exposure to radiation of 2537 Å.; this led to the discovery of new instances of photosensitization, and to attempts to elucidate the mechanism of such processes. When hydrogen is mixed with oxygen, carbon monoxide, ethylene, nitrous oxide or nitrogen, and the gases saturated with mercury vapor at ordinary temperatures, reaction occurs on exposure to radiation from a mercury vapor lamp (H. S. Taylor, 1925). The processes are generally very complicated, but some attempts have been made to interpret them.

The photosensitized reaction between hydrogen and oxygen in the presence of mercury vapor leads to the formation of water and hydrogen peroxide; under certain conditions the latter predominates. The first stage is undoubtedly the formation of hydrogen atoms by collision between the excited mercury atom and a hydrogen molecule (p. 1157), and this may be followed by reactions such as



with $\text{HO}_2 + \text{H}_2 = \text{H}_2\text{O}_2 + \text{H}$, as a further possibility; this would imply a chain mechanism. The hydrogen peroxide may be isolated as such, or it may decompose to yield water and oxygen.

The quantum yield for the reaction between hydrogen and carbon monoxide is approximately two, and the products are formaldehyde and glyoxal in similar amounts; the mechanism suggested is



followed by either



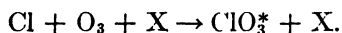
but the process may actually be more complex. Complications undoubtedly occur in the photosensitized reduction of ethylene by hydrogen in the presence of mercury. When neither reactant is in excess the gaseous product is mainly butane, together with about 20 per cent of ethane; at high temperatures some methane is also formed. In the absence of hydrogen, ethylene is polymerized to a considerable extent under the same conditions.

Mercury is able to photosensitize the decomposition of ammonia by 2537 Å. radiation, the products of the first stage being apparently the same as those for the direct photolysis (p. 1176), i.e., $\text{NH}_2 + \text{H}$, although the photosensitized rate is about 200 times as great, at the same wave length. Phosphine undergoes similar decomposition. The processes are much more complex than appears at first sight.

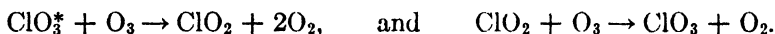
In addition to the cases already mentioned, the presence of a little mercury vapor and exposure to the 2537 Å. radiation leads to the ozonization of oxygen, with a quantum yield of about seven; the process is consequently a chain reaction. Water vapor and a number of organic compounds, e.g., ethyl alcohol and acetone, can be decomposed photochemically in a similar manner.

Although the energy of the resonance radiation of cadmium is insuffi-

cient to cause dissociation of hydrogen (p. 1157), it is nevertheless large enough to break C—C bonds. In recent years, therefore, reactions photosensitized by cadmium vapor have attracted some attention (E. W. R. Steacie, 1939 *et seq.*). The halogens function as photosensitizers in a number of processes (cf. p. 1180). A reaction which has been studied is the decomposition of ozone, in the presence of chlorine, by means of ultra-violet light; this radiation is absorbed by chlorine and dissociates it, but affects ozone to a slight extent only. In the mixture of gases, however, the ozone is readily decomposed, although the chlorine is almost unchanged at the end; the yield is approximately two molecules of ozone decomposed per quantum absorbed. The rate of reaction is independent of the ozone concentration, and is proportional to the intensity of the absorbed light. The process is unexpectedly complicated, involving a chain mechanism and surface reaction. Put in the simplest form, the various stages are believed to be, first, the direct formation of chlorine atoms by absorption of radiation of less than 4785 Å. wave length, followed by



There are reasons for believing that ClO_3^* is formed, since the acids HClO_3 and HClO_4 have been detected when a little moisture is present, and the polymer Cl_2O_6 has been isolated and identified. The excited ClO_3^* radical may then be adsorbed on the walls of the vessel where it partly decomposes to yield chlorine and oxygen, and partly polymerizes to Cl_2O_6 , which also decomposes to give the same products. In addition the ClO_3^* will react in the gas phase; thus,

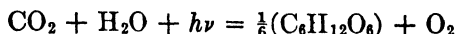


The reaction between ClO_3^* and ozone requires an energy of activation of about 12 kcal., so the process would not be very rapid at room temperature unless one of the substances was in an excited state; it is for this reason that reaction with ClO_3^* is postulated. Many other mechanisms have been proposed from time to time, but the one described appears to cover many, but not all, of the facts (A. J. Allmand, 1932). It is possible that ClO , which appears to be involved in the thermal chlorine-ozone reaction, may also take part in the photochemical process under consideration. The decomposition of ozone is also sensitized by bromine; the quantum yield is apparently 30, and it is evident that the process must be quite different from that with chlorine.³⁷

Examples of photosensitization in solution are known and are of practical importance; mention has been made of the use of uranyl salts to sensitize the decomposition of oxalic acid by light of short wave length (p. 1162). In the absence of the uranyl compound the quantum yield is very low, generally less than 0.01, but for the sensitized reaction it is 0.5 or more; the mechanism of the decomposition is not well understood. Uranyl salts also sensitize the photolysis of formic acid solutions, and ferric salts play the same part in the reaction occurring in the Eder acti-

nometer (p. 1162), and in the oxidation of potassium ferrocyanide in alkaline solution.

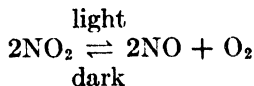
The most familiar instance of photosensitization is the action of chlorophyll in permitting carbon dioxide and water to react in sunlight, a process resulting eventually in the formation of carbohydrates. The reaction is remarkable in the respect that



requires 112 kcal. per mole of carbon dioxide, yet in the living plant, in the presence of chlorophyll, it can take place even in red light of 7000 Å. wave length which corresponds to only 38 kcal. per mole. The process is obviously a very complex one, and the maximum possible quantum yield for red light cannot exceed 38/112, i.e., 0.34, molecule of carbon dioxide reacting per quantum absorbed. According to O. Warburg (1923) the quantum yield observed with a species of algae, *Chlorella vulgaris*, was about 0.23 for light of wave lengths varying from 4360 to 6600 Å. Some doubt has, however, been cast on this result, and it is now believed that the efficiency of photosynthesis is not more than 0.05 to 0.1 molecule of carbon dioxide per quantum (F. Daniels, *et al.*, 1939). The sensitizing action of chlorophyll is certainly remarkable, but it is not confined to the photosynthesis of carbohydrates, for this substance, in common with a number of dyestuffs, is able to bring about oxidation reactions in the presence of light.³⁸

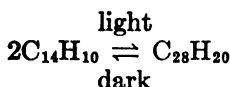
Photochemical Equilibrium.—If either or both the reactions in a reversible chemical change are sensitive to light, a **photochemical equilibrium**, or **photostationary state**, will be set up when the rates of direct and reverse reactions are equal. The position of the equilibrium will, however, be different from the normal thermal equilibrium, since energy is being added to the system by the absorbed radiation; the photochemical equilibrium state will, therefore, depend on the intensity and wave length of the absorbed light. After the equilibrium condition is reached the absorbed light can produce no further chemical change, and so it is converted into heat. When the source of radiation is removed the system will tend to change to the true, thermodynamic equilibrium for the given temperature and pressure; if the temperature is low, however, an unstable condition may persist for a considerable time because of the slow rate of attainment of equilibrium. A number of cases of the photochemical stationary state have been studied; they fall into two categories depending on whether light affects one or both of the reactions concerned.

When nitrogen dioxide vapor is exposed to light of wave length less than 3700 Å. it decomposes into nitric oxide and oxygen; the combination of these products is a simple thermal, or dark, reaction, and so the equilibrium



is set up (R. G. W. Norrish, 1927). The pressure of the gas rises at first, due to decomposition of the dioxide, and then becomes constant when the stationary state is attained; part of the increase of pressure is, of course, the result of an increase of temperature. In calculating the quantum yield of the photoprocess allowance must be made for the reverse reaction, and in this way a value of almost exactly two was found for wave lengths shorter than 3700 Å.; the *apparent* yield is less than this value, and is decreased by addition of oxygen or nitric oxide to the reacting system.

Another illustration of a photostationary state involving one light-sensitive reaction is the dimerization of anthracene in solution (p. 1168); the depolymerization process is a thermal reaction, and so the equilibrium



is eventually set up. At constant temperature the rate of the direct reaction is proportional to the absorbed light ($I_{\text{abs.}}$) whereas the reverse reaction, assuming it to be of the first order, will be proportional to the concentration of dianthracene, c_{A_2} ; when the photostationary state is attained,

$$kI_{\text{abs.}} = k'c_{A_2}; \quad \therefore c_{A_2} = kI_{\text{abs.}}/k', \quad (214)^*$$

where k and k' are constants for the light and dark reactions, respectively. Provided $I_{\text{abs.}}$ remains constant, the equilibrium concentration of dianthracene should thus be independent of that of anthracene, as is the case for initial concentrations of the latter exceeding 0.16 mole per liter. This result emphasizes the fact that the ordinary law of thermal equilibrium, leading to the usual equilibrium constant, is not applicable directly to photostationary states. The variation of the dianthracene concentration with temperature should depend on the relative changes in k and k' ; the temperature coefficients are 1.1 and 2.8 respectively, for a rise of 10°, and so the quantity c_{A_2} at equilibrium should change $1.1/2.8 = 0.39$ -fold for every 10° rise of temperature. The experimental value was found to be 0.34 (R. Luther and F. Weigert, 1905).

Both direct and reverse reactions in the system



take place with appreciable velocity when exposed to ultraviolet radiation, although neither occurs to any extent in the dark at ordinary temperatures. The photochemical reactions have been studied by A. Coehn and H. Becker (1907-09) and the equilibrium concentrations of the various reactants determined; these depend on the experimental conditions, e.g., wave length and intensity of absorbed light, but they differ completely from those at the position of thermal equilibrium. For

* In this expression $I_{\text{abs.}}$ represents the *net* absorption, after allowing for light emitted as fluorescence (see p. 1169).

example, at 45° c. the normal decomposition of sulfur trioxide occurs to a very small extent only, but under the experimental conditions the photostationary state was found to correspond to 35 per cent decomposition. The state of photochemical equilibrium is almost independent of temperature from 50° to 800° c.; since the homogeneous thermal reaction, in the absence of a catalyst, occurs at a negligible rate over this range it appears that the direct and reverse photochemical reactions have the same temperature coefficient. It was thought at one time that the law of thermal equilibrium applied to the photostationary state, and that the concentrations satisfied the requirements of an equilibrium constant; this is, however, improbable for it is very unusual to find the rate of a photochemical reaction to be directly proportional to the concentration of the reactants, as the existence of an equilibrium constant would imply.

A number of other reversible processes, in which one or both reactions are photochemical, have been studied, but as they present no special features it is only necessary to mention some of them: for example, $\text{CO} + \text{Cl}_2 \rightleftharpoons \text{COCl}_2$; $2\text{CO} + \text{O}_2 \rightleftharpoons 2\text{CO}_2$; $2\text{HX} \rightleftharpoons \text{H}_2 + \text{X}_2$, where X is a halogen; $\text{SO}_2 + \text{Cl}_2 \rightleftharpoons \text{SO}_2\text{Cl}_2$; and fumaric acid \rightleftharpoons maleic acid (p. 1180).³⁹

Attempts have been made to explain the phenomena of vision by postulating a photostationary state. The light sensitive substance in the eye, known as "visual purple," is bleached, forming "visual yellow," on exposure to light, but it is regenerated in a thermal change, with the result that an equilibrium is established. In the dark the visual purple accumulates, and the eye becomes especially sensitive; exposure to light then results in the phenomenon of dazzle. Although the simple purple \rightleftharpoons yellow equilibrium probably constitutes the basis of vision, there are undoubtedly many other factors involved.

Influence of Temperature.—The temperature coefficients of photochemical processes are on the whole much smaller than for thermal reactions (E. Goldberg, 1903); values between 1.0 and 1.1 are common for a 10° rise of temperature, and those above 1.5 are rare, although the reaction between oxalate ions and iodine in solution has a temperature coefficient as high as 3.2. In studying the influence of temperature on photo-processes it is best to examine the effect on the quantum yield, since the rate of reaction at constant intensity of the incident light may vary because of the difference in absorption at different temperatures. The quantum yield is, however, free from any such complications.

The problem may be considered from two points of view, according as the yield is determined by the primary light-absorption process or by subsequent chemical reactions. For processes obeying the Stark-Einstein photochemical equivalence law, that is, with quantum yields of 1, 2 or perhaps 3, the influence of temperature must be small. The yield is virtually determined by the light-absorbing stage, and since every quantum absorbed ultimately leads to the reaction of one or two molecules, temperature can have no effect unless it brings about a fundamental change. In cases of this kind the temperature coefficient will be unity,

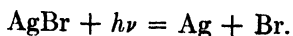
and the closer the actual behavior approaches to the ideal, e.g., in the decomposition of hydrogen iodide, the nearer will be the coefficient to this value. If the quantum yield is low, because of loss of energy by the excited molecules, e.g., in the polymerization of anthracene, or because of recombination of the initially formed products, e.g., photolysis of ammonia, the influence of temperature will depend on the relative increases in the rates of these processes and those whereby the excited molecules or their products take part in further reactions. In any case, the temperature coefficient is not likely to be greatly different from unity.

When the quantum yield is high, because a chain mechanism is operative, the various chemical reactions involved in the chain must have low energies of activation, otherwise they would not occur so readily at ordinary temperatures (cf. p. 1091); a small activation energy means a small temperature coefficient, which again will not be much larger than unity. A chain mechanism associated with a relatively low quantum yield, e.g., in the hydrogen-bromine reaction, implies that an important intermediate stage is slow and involves a relatively large amount of activation; the temperature coefficient for such a process may then be in excess of unity, e.g., 1.5 for the reaction mentioned. Another possibility is that a thermal equilibrium, varying with temperature, may influence the photochemical process.

The foregoing arguments have been based on the assumption that the primary process is the same for radiation of a given wave length at all temperatures, but this may not be true. At higher temperatures a larger proportion of the molecules are in higher vibrational levels, and consequently the energy required to bring about dissociation will be smaller than at lower temperatures. As a result the region of continuous or of diffuse absorption of light will commence at longer wave lengths as the temperature is raised. It is possible, therefore, that radiation of a given frequency will be able to cause dissociation (or predissociation) at a higher but not at a lower temperature; in circumstances of this kind quite large temperature coefficients might be observed.

Photoreactions in Solids: Silver Halides.—Although many solids are sensitive to light, only two types of reaction involving solids appear to have been investigated: one is concerned with the decomposition of silver halides on the photographic plate, and the other with the isomeric change of *o*-nitrobenzaldehyde to *o*-nitrosobenzoic acid. Pure silver halides are said to be almost unaffected by light, but small quantities of allyl isothiocyanate, and possibly other substances, present in the "emulsion" used for coating the plates, act as sensitizers. As is well known, the sensitivity depends on the size of the silver halide grains and other factors, not all clearly understood, involved in the preparation of the photographic "emulsions." A number of studies have been made of the primary light-absorption process; it involves decomposition of the silver halide into silver and halogen, and by estimating the amount of the latter the quantum yield has been found to be very close to unity in

several instances (J. Eggert and W. Noddack, 1921–29), i.e.,



In the crystal the silver bromide actually consists of silver and bromine ions; hence the effect of radiation is to remove an electron from the bromine ion and transfer it to the silver ion. The process is, therefore, in a sense analogous to the photoelectric effect with the exception that the electron is not expelled. When exposed to a strong source of light the silver produced by the decomposition of the halide is evident by the discoloration, but in the ordinary way a photographic plate shows no evidence of exposure. Nevertheless, a "latent image" is present which is developed by means of a suitable reducing agent; it is not only the silver produced by direct photochemical action that becomes visible, for this silver nucleus apparently catalyzes, possibly by initiating chains, the reduction of a large number, approximately 300, of surrounding silver bromide molecules. The extent of this reduction depends on the intensity and wave length of the original light, thus giving the familiar variation of "density" in the negative. The exact mechanism of the process of development and the function of the latent image are still uncertain, and a further consideration of the subject is outside the scope of this book. Mention may be made, however, of the use of sensitizers in photography; it was discovered by H. Vogel as far back as 1873 that the addition of certain dyestuffs to the photographic "emulsion" made the plates sensitive to longer wave lengths than usual. This fact has been utilized, especially in recent years, to sensitize photographic plates for infrared rays (cf. p. 583). It appears that the dyestuffs are actually adsorbed by the silver halide, and it is possible that their action may depend on the formation of a definite compound. The reverse phenomenon of **desensitization** by means of dyestuffs is also known; photographic plates treated with these substances are no longer affected by light of relatively short wave length.

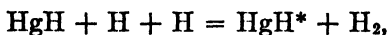
Isomeric Change.—The isomerization of *o*-nitrobenzaldehyde to *o*-nitrosobenzoic acid is a photoreaction of unusual interest as it takes place with the same efficiency, viz., 0.5, in the solid state and in solution in acetone or ligroin; the effective radiation is in the range from 4360 to 3130 Å. The analogous change, with the same efficiency, also occurs with 2:4-dinitrobenzaldehyde. A variety of suggestions have been made to explain the abnormal quantum yield 0.5; it cannot be the result of chance as it is a constant under such different conditions. According to K. G. Zimmer (1933) the primary photochemical stage is the splitting off of an oxygen atom from the nitro-group; if the probability of its attaching itself to the —CHO group is equal to that of its returning to its original position, the yield will be 0.5 molecule per quantum. This view has been criticized for several reasons: first, no evolution of oxygen has been detected; second, it is not certain that the energy of the radiation is suffi-

cient to break the oxygen-nitrogen linkage; and third, *m*- and *p*-nitrobenzaldehydes do not undergo the isomeric change although it might be expected from the proposed mechanism. Another theory is that after excitation a definite time elapses before the molecule undergoes isomeric change; if this time is approximately equal to the average life of the excited molecule before losing its energy, the quantum yield will be of the observed order (P. A. Leighton, 1934). The subject is interesting and merits further investigation.⁴⁰

CHEMILUMINESCENCE

The term **luminescence** is used to describe the emission of radiation in excess of the "black-body" radiation for the same temperature; for example, below about 500° c. a black body does not emit visible or ultra-violet radiation to any appreciable extent, and so if by some means a system were made to emit such radiations, it would be described as luminescent. Various types of luminescence are known, of which **chemiluminescence**, signifying the emission of light as the result of a chemical reaction at a temperature when light rays are normally not to be expected, is probably the most important. In a sense chemiluminescence may be regarded as the reverse of a photochemical reaction, and hence some chemiluminescent processes will be described here.

Combination of Atoms.—When a stream of atomic hydrogen falls on a surface of liquid mercury a blue luminescence appears, the spectrum consisting of the resonance line 2537 Å., and a band system from 4500 to 3250 Å. due to the mercury hydride HgH (K. Bonhoeffer, 1925). It appears that cold liquid mercury must be used since the vapor gives no luminescence, nor does the liquid if its temperature is raised. The formation of HgH is probably a heterogeneous process, and when the hydride enters the vapor phase it will have little excess energy. Excitation of the mercury hydride to a higher electronic level, so that on returning to the ground state it can emit the radiation mentioned above, may then occur as a result of the reaction



the energy evolved in the combination of the hydrogen atoms, approximately 100 kcal., being sufficient for the purpose. The emission of the 2537 Å. line by ³P₁ mercury atoms cannot be due to direct excitation by the heat of combination of hydrogen atoms, as 112 kcal. are required (p. 1157); a probable explanation is that the excited mercury atom is formed in a collision between a normal and an excited HgH molecule, viz.,



The combination of hydrogen atoms in the presence of sodium vapor results in the emission of the *D*-lines of the sodium spectrum; the energy of the excited level is only 48 kcal. per g. atom, which is less than that

given out in the formation of molecular hydrogen. The "glow" of active nitrogen belongs to the same type of reaction. When two nitrogen atoms combine a three-body collision is necessary, and if a nitrogen molecule is involved (see p. 1065) it is raised to a high electronic level; on returning to the ground state the characteristic glow is emitted.

Alkali Metal Vapors and Halogens.—Alkali metal vapors react with halogens and with organic halides at low pressures, and as a result a luminescence consisting of the spectrum of the metal is produced (F. Haber and W. Zisch, 1922). The reactions have been studied by M. Polanyi *et al.* (1928 *et seq.*), at pressures of 10^{-2} to 10^{-3} mm. of mercury, when the so-called "highly dilute" flames are obtained (p. 1097). In order to account for the observations it is necessary to postulate a somewhat complex mechanism of three stages, in the second of which, i.e., between Na_2 molecules and Cl atoms, sufficient energy is obtained by the NaCl molecule for it to excite a sodium atom on collision; thus,



On returning to its normal state the excited atom emits its characteristic spectrum. Similar luminescence occurs in the reaction between sodium or potassium vapor and mercuric chloride vapor at very low pressures.

Oxidation Reactions.—The glow of phosphorus and of its trioxide provides the most familiar example of chemiluminescence; the light is probably emitted by an oxide in an electronically excited state. The actual reaction is believed to be a chain process, for in this way the upper and lower pressure limits of oxygen at which the glow occurs can be explained (cf. p. 1083). Organic compounds, e.g., ethylene, acetone and carbon disulfide, which inhibit the glow, presumably act as chain breakers.

Grignard reagents are an important source of chemiluminescence. The reaction with chloropicrin is accompanied by light (E. Wedekind, 1906), and so also is the oxidation by air or oxygen of Grignard compounds in the solid state or in ethereal solution (I. Lifschitz and O. E. Kalberer, 1922; W. V. Evans, 1923–26). The luminescence is greenish-blue and generally consists of a single broad band. Only Grignard reagents in which the magnesium atom is attached to an unsaturated carbon atom, e.g., in a benzene ring, are chemiluminescent in solution, but in the solid state both aliphatic and aromatic compounds exhibit the phenomenon. It is of interest that no luminescence accompanies oxidation of Grignard compounds by hydrogen peroxide.

The oxidation of alkaline aqueous solutions of 5-aminophthalic hydrazide, or of other cyclic hydrazides, is accompanied by luminescence; various oxidizing agents may be used, hydrogen peroxide giving good results. The addition of a substance which accelerates the decomposition of the peroxide, e.g., potassium ferricyanide, sodium hypochlorite or colloidal platinum, increases the rate of oxidation and the intensity of the light emitted (H. O. Albrecht, 1928).

Treatment of calcium silicate with aqueous hydrochloric acid gives

rise to a number of complex solid products varying in empirical formula from $\text{H}_2\text{Si}_2\text{O}$ (siloxen) to that of silicic acid, H_2SiO_3 . These substances may be oxidized by air or oxygen or by an oxidizing agent, and the reaction is accompanied by appreciable luminescence (H. Kautsky *et al.*, 1922–26). If the oxidation is carried out, e.g., by permanganate, in acid solution in the presence of certain dyestuffs which are adsorbed by the silicon compound, e.g., rhodamine B, a strong red fluorescence, characteristic of the dye, is observed. The energy produced in the oxidator is transferred to the adsorbed dyestuff which is thus raised to an excited level; the return to the normal state is accompanied by the production of the fluorescence. This is an example of sensitized chemiluminescence.

In conclusion, reference may be made to the cold light produced by certain living organisms: this accompanies the oxidation by atmospheric oxygen of a protein derivative luciferin, in the presence of an enzyme known as luciferase (E. N. Harvey, 1915 *et seq.*).⁴¹

REFERENCES

- Bell, R. P. "Acid-Base Catalysis"
 Daniels, F. "Chemical Kinetics."
 Glasstone, S., Laidler, K. J. and Eyring, H. "The Theory of Rate Processes."
 Griffith, R. O. and McKeown, A. "Photoprocesses, etc."
 Hinshelwood, C. N. "Kinetics of Chemical Change."
 Kassel, L. "Kinetics of Homogeneous Gas Reactions"
 Kistiakowsky, G. B. "Photochemical Processes."
 Mellor, J. "Chemical Statics and Dynamics."
 Moelwyn-Hughes, E. A. "Kinetics of Reactions in Solution."
 Noyes, W. A. and Leighton, P. A. "The Photochemistry of Gases."
 Pease, R. N. "Equilibrium and Kinetics of Gas Reactions."
 Rideal, E. K. and Taylor, H. S. "Catalysis in Theory and Practice."
 Rollefson, G. K. and Burton, M. "Photochemistry."
 Schwab, G. M. "Catalysis" (Trans. by H. S. Taylor and R. Spence).
 Semenov, N. "Chemical Kinetics and Chain Reactions."
 Style, D. W. G. "Photochemistry."
 Taylor, H. S. "Treatise on Physical Chemistry," 2nd ed., Chap XIV (F. O. Rice and H. C. Urey); Chap. XV and XVIII (H. S. Taylor).
1. Daniels and Johnston, *J. Am. Chem. Soc.*, **43**, 53 (1921); Hinshelwood and Hutchison, *Proc. Roy. Soc.*, **111A**, 245 (1926); Ramsperger, *J. Am. Chem. Soc.*, **49**, 912, 1495 (1927); H. A. Taylor, *J. Phys. Chem.*, **34**, 2761 (1930); **35**, 2658 (1931); Smith, *J. Am. Chem. Soc.*, **49**, 43 (1927).
 2. Eyring and Daniels, *ibid.*, **52**, 1472 (1930); Hinshelwood and Moelwyn-Hughes, *Proc. Roy. Soc.*, **131A**, 186 (1931); van't Hoff, "Studies in Chemical Dynamics."
 3. Reicher, *Annalen*, **238**, 276 (1887); Walker, *Proc. Roy. Soc.*, **78A**, 157 (1906); Trautz, *Z. anorg. Chem.*, **88**, 285 (1914); **102**, 149 (1918); Bodenstein and Lindner, *Z. phys. Chem.*, **100**, 68 (1922); Hinshelwood, *et al.*, *J. Chem. Soc.*, 730 (1926); 378 (1936); Rice, *J. Chem. Phys.*, **4**, 53 (1936); Gorin, *J. Am. Chem. Soc.*, **58**, 1787 (1936); von Kiss, *Z. anorg. Chem.*, **191**, 289 (1930); *Rec. trav. chim.*, **52**, 289 (1933); Smith, *J. Am. Chem. Soc.*, **65**, 74 (1943); Weiss, *J. Chem. Soc.*, 309 (1944).
 5. Polanyi, *Z. Phys.*, **1**, 337 (1920); Herzfeld, *ibid.*, **8**, 132 (1922); Born and Franck, *Ann. Phys.*, **76**, 225 (1925); Willey, *J. Chem. Soc.*, 2831 (1927); 228 (1928).
 5. Lowry, *ibid.*, **73**, 211 (1899); Walker and Appleyard, *ibid.*, **67**, 193 (1896); Bodenstein, *Z. phys. Chem.*, **29**, 295 (1899); Kistiakowsky, *J. Am. Chem. Soc.*, **50**, 2315 (1928); Taylor and Crist, *ibid.*, **63**, 1377 (1941).

6. Burgess and Chapman, *J. Chem. Soc.*, **89**, 1399 (1906); Landolt, *Ber.*, **19**, 1317 (1886); Eggert, *Z. Elek.*, **23**, 8 (1917); Skrabal, *ibid.*, **28**, 224 (1922); Roman-Levinson, *ibid.*, **34**, 333 (1928); Bray, *J. Am. Chem. Soc.*, **43**, 1262 (1921).
7. Bodenstein, *et al.*, *Z. phys. Chem.*, **57**, 168 (1906); **114**, 208 (1924); Herzfeld, *Z. Elek.*, **25**, 301 (1919); Polanyi, *ibid.*, **26**, 49 (1920).
8. Nernst, *ibid.*, **24**, 335 (1918); Polanyi, *Trans. Farad. Soc.*, **24**, 606 (1928); Hinshelwood, *et al.*, *Proc. Roy. Soc.*, **118A**, 170 (1928); **119A**, 591; **122A**, 610 (1929); **138A**, 311 (1932); **180A**, 237 (1942); Alyea and Haber, *Z. phys. Chem.*, **B10**, 193 (1930); Chariton and Walta, *Z. Phys.*, **39**, 547 (1926); Semenov, *ibid.*, **46**, 109 (1927); *op. cit.*; Rice and Herzfeld, *J. Am. Chem. Soc.*, **56**, 284 (1934); *J. Chem. Phys.*, **7**, 671 (1939); Hinshelwood, *J. Chem. Soc.*, 818 (1936); 1568 (1937); Pease, *op. cit.*, Chap. VII.
9. Wegscheider, *Z. phys. Chem.*, **30**, 593 (1899); **34**, 290 (1900); Hollemann, *ibid.*, **31**, 79 (1899); *Rec. trav. chim.*, **18**, 267 (1899); **19**, 79 (1900); Bredig and Balcom, *Ber.*, **41**, 740 (1908).
10. Arrhenius, *Z. phys. Chem.*, **4**, 226 (1889); see *Trans. Farad. Soc.*, **33**, 72 (1937); Mayer, *J. Am. Chem. Soc.*, **49**, 3033 (1927).
11. London, *Z. Elek.*, **35**, 552 (1929); Eyring and Polanyi, *Z. phys. Chem.*, **B12**, 279 (1931); Eyring, *J. Am. Chem. Soc.*, **53**, 2537 (1931); *Chem. Rev.*, **10**, 103 (1932); *Trans. Farad. Soc.*, **34**, 3 (1938); Glasstone, Laidler and Eyring, *op. cit.*, Chap. III.
12. Lewis, *J. Chem. Soc.*, **113**, 471 (1918); Hinshelwood, *op. cit.*
13. Polanyi, *et al.*, *Z. phys. Chem.*, **B1**, 3, 21, 30, 62 (1928); *ibid.*, **B11**, 97 (1930); **B19**, 139 (1932); *Z. Phys.*, **47**, 379 (1928); "Atomic Reactions"; Bawn, *Ann. Rep. Chem. Soc.*, **39**, 36 (1943).
14. Hinshelwood and Moelwyn-Hughes, *Proc. Roy. Soc.*, **131A**, 177; **134A**, 211 (1931); *J. Chem. Soc.*, 230 (1932).
15. Eyring, *J. Chem. Phys.*, **3**, 107, 492 (1935); *Chem. Rev.*, **17**, 65 (1935); *Trans. Farad. Soc.*, **34**, 41 (1938); Polanyi and Evans, *ibid.*, **31**, 875 (1935); for complete review, see Glasstone, Laidler and Eyring, *op. cit.*
16. Farkas and Farkas, *Proc. Roy. Soc.*, **152A**, 124, 152 (1935).
17. Christiansen and Kramers, *Z. phys. Chem.*, **104**, 451 (1923); Lindemann, *Trans. Farad. Soc.*, **17**, 598 (1922); Hinshelwood, *Proc. Roy. Soc.*, **113A**, 230; **114A**, 84 (1927); Fowler and Rideal, *ibid.*, **113A**, 570 (1927); Rice and Ramsperger, *J. Am. Chem. Soc.*, **49**, 1617 (1927); **50**, 617 (1928); Kassel, *J. Phys. Chem.*, **32**, 225 (1928); for review, see Pease, *op. cit.*, Chap. VI.
18. Grant and Hinshelwood, *J. Chem. Soc.*, 258 (1933); Moelwyn-Hughes, *Chem. Rev.*, **10**, 241 (1932); *op. cit.*; Glasstone, *et al.*, *op. cit.*, Chap. VIII; Bell, *J. Chem. Soc.*, 629 (1943).
19. Menshutkin, *Z. phys. Chem.*, **1**, 629 (1887); **6**, 41 (1889); Patterson and Montgomerie, *J. Chem. Soc.*, **101**, 26, 2100 (1912); Soper, *et al.*, *ibid.*, 1873 (1929); 2297 (1931); Glasstone, *ibid.*, 723 (1936); Wood and Dickinson, *J. Am. Chem. Soc.*, **61**, 3259 (1939).
20. Brønsted, *Z. phys. Chem.*, **102**, 169 (1922); **115**, 337 (1925); Bjerrum, *ibid.*, **108**, 82 (1924); **118**, 251 (1925); for reviews, see *Ann. Rep. Chem. Soc.*, **23**, 30 (1926); **24**, 330 (1927); **31**, 67 (1934); La Mer, *Chem. Rev.*, **10**, 179 (1932).
21. Langmuir, *J. Am. Chem. Soc.*, **38**, 2221 (1916); **40**, 1361 (1918); *Trans. Farad. Soc.*, **17**, 607, 621 (1922); Hinshelwood, *et al.*, *J. Chem. Soc.*, **123**, 1014 (1923); **125**, 393 (1924); **127**, 327, 1116, 1552, 2896 (1925); Kunsman, *J. Am. Chem. Soc.*, **50**, 2100 (1928); Schwab and Pietsch, *Z. phys. Chem.*, **B1**, 385 (1928); Bachmann and Taylor, *J. Phys. Chem.*, **33**, 447 (1929); Bodenstein and Fink, *Z. phys. Chem.*, **60**, 1 (1907); Glasstone, Laidler and Eyring, *op. cit.*, Chap. VII.
22. Braune and Knoke, *Z. phys. Chem.*, **135**, 49 (1928); Rodebush and Michalek, *J. Am. Chem. Soc.*, **51**, 748 (1929); Gucker and Munch, *ibid.*, **59**, 1275 (1937); Briner, *J. Chim. phys.*, **23**, 848 (1926); Dixon, *Phil. Trans. Roy. Soc.*, **175**, 617 (1884); Weston, *Proc. Roy. Soc.*, **109A**, 176, 523 (1925); Hinshelwood, *et al.*, *ibid.*, **124A**, 219 (1929); *Trans. Farad. Soc.*, **24**, 559 (1928); *J. Chem. Soc.*, 464 (1940); von Elbe

- and Lewis, *J. Am. Chem. Soc.*, **59**, 2022, 2025 (1937); Crist and Wertz, *J. Chem. Phys.*, **7**, 719 (1939).
23. Arrhenius, *Z. phys. Chem.*, **31**, 197 (1899); Senter, *J. Chem. Soc.*, **91**, 460 (1907); Taylor, *Z. Elek.*, **20**, 202 (1914); *J. Am. Chem. Soc.*, **37**, 551 (1915); Dawson, *J. Chem. Soc.*, **107**, 1426 (1915).
24. Brønsted, *et al.*, *Z. phys. Chem.*, **108**, 185 (1924); *J. Am. Chem. Soc.*, **49**, 2554 (1927); **52**, 1394 (1930); *Trans. Farad. Soc.*, **24**, 630 (1928); **25**, 59 (1929); *Chem. Rev.*, **5**, 231 (1928); Dawson, *J. Chem. Soc.*, **2282**, 2872, 3166 (1926); *Trans. Farad. Soc.*, **24**, 641 (1928); Harned, *ibid.*, p. 666; Lowry, *ibid.*, p. 545; *J. Chem. Soc.*, **1371**, 1385, 2883 (1925); 1938 (1926); Kilpatrick and Kilpatrick, *Chem. Rev.*, **10**, 213 (1932); Bell, *et al.*, *Proc. Roy. Soc.*, **143A**, 377 (1934); **151A**, 211 (1935); **176A**, 88, 114 (1940); *J. Chem. Soc.*, **382**, 1861 (1938); 1096, 1774, 1777 (1939); 1202 (1940); *op. cit.*; Pflugger, *J. Am. Chem. Soc.*, **60**, 1513 (1938); King, *ibid.*, **62**, 379 (1940); Olson and Tong, *ibid.*, **66**, 1555 (1944).
25. Bray, *Chem. Rev.*, **10**, 161 (1932).
26. Bigelow, *Z. phys. Chem.*, **26**, 493 (1898); Titoff, *ibid.*, **45**, 641 (1903); Christiansen, *J. Phys. Chem.*, **28**, 145 (1924); Bäckström, *J. Am. Chem. Soc.*, **49**, 1460 (1927); **51**, 90 (1929); Wiig, *ibid.*, **52**, 4742 (1930); Fuller and Crist, *ibid.*, **63**, 1644 (1941).
27. Pease, *J. Am. Chem. Soc.*, **45**, 1196, 2235, 2296 (1923); Hinshelwood and Prichard, *J. Chem. Soc.*, **127**, 806 (1925); Taylor, *Proc. Roy. Soc.*, **108A**, 105 (1925); **113A**, 77 (1926); *Chem. Rev.*, **9**, 1 (1931); Burk, *J. Phys. Chem.*, **30**, 1134 (1926); Balandin, *Z. phys. Chem.*, **B2**, 289; **B3**, 167 (1929); Palmer and Constable, *Proc. Roy. Soc.*, **107A**, 255, 270 (1925); Wilkins, *J. Chem. Soc.*, **330** (1931); Beeck, Smith and Wheeler, *Proc. Roy. Soc.*, **177A**, 62 (1940); Russell and Lobenstein, *J. Am. Chem. Soc.*, **62**, 2573 (1940); Emmett and Skau, *ibid.*, **65**, 1029 (1943); Pease, *op. cit.*, Chap. IX; see also, refs. (21).
28. Cario and Franck, *Z. Phys.*, **11**, 161 (1922); **17**, 202 (1923); Taylor and Marshall, *J. Phys. Chem.*, **29**, 1140 (1925); Hirschlaff, "Fluorescence and Phosphorescence."
29. Lewis, *et al.*, *J. Am. Chem. Soc.*, **63**, 3005 (1941); **64**, 1774 (1942); **66**, 2100 (1944).
30. Einstein, *Ann. Phys.*, **37**, 832; **38**, 881 (1912); *J. de Phys.*, **8**, 277 (1913); Stark, *Phys. Z.*, **9**, 889, 894 (1908); Bodenstein, *Z. phys. Chem.*, **85**, 329 (1913); for discussion, see Allmand, *Ann. Rep. Chem. Soc.*, **22**, 333 (1925).
31. Experimental methods are reviewed by Forbes, *J. Phys. Chem.*, **32**, 485 (1928); Daniels, *ibid.*, **42**, 701 (1938).
32. For complete summaries, see Daniels, *ibid.*, p. 713; Noyes and Leighton, *op. cit.*
33. Franck, *Trans. Farad. Soc.*, **21**, 536 (1925); Henri, *Compt. rend.*, **179**, 1156 (1924); *Trans. Farad. Soc.*, **25**, 765 (1929); Herzberg, *Z. Phys.*, **61**, 604 (1930); Burton and Rollefson, *J. Chem. Phys.*, **6**, 416, 674 (1938).
34. For summaries of the large amount of work done on the hydrogen-chlorine reaction, see Griffith and McKeown, *op. cit.*; *Ann. Rep. Chem. Soc.*, **28**, 46 (1931); **30**, 49 (1933); **34**, 62 (1937); Ritchie and Taylor, *Proc. Roy. Soc.*, **180A**, 423 (1942); see also, Bodenstein, *Z. phys. Chem.*, **B48**, 239, 268 (1941).
35. For reviews with references, see Taylor, *J. Phys. Chem.*, **42**, 783 (1938); Leighton, *ibid.*, p. 749; see also Norrish, *et al.*, *J. Chem. Soc.*, 1521, 1531, 1544 (1938); Grahame and Rollefson, *J. Chem. Phys.*, **7**, 775 (1939); **8**, 98 (1940); Burton, *et al.*, **7**, 1075, 1080 (1939); Howland and Noyes, *J. Am. Chem. Soc.*, **66**, 974 (1944).
36. Olson, *et al.*, *Trans. Farad. Soc.*, **27**, 69 (1931); *J. Am. Chem. Soc.*, **55**, 1410 (1933); **56**, 1320 (1934); Vaidya, *Proc. Roy. Soc.*, **129A**, 299 (1930); Brakula, *Z. phys. Chem.*, **B25**, 90 (1934); Atwood and Rollefson, *J. Chem. Phys.*, **9**, 506 (1941).
37. Taylor and Marshall, *ref. (28)*; Taylor, *et al.*, *J. Chem. Phys.*, **5**, 203, 212 (1937); Melville, *et al.*, *Proc. Roy. Soc.*, **157A**, 621 (1936); **160A**, 384, 406 (1937); Allmand, *J. Chem. Soc.*, 599 (1932); Heidt, Kistiakowsky and Forbes, *J. Am. Chem. Soc.*, **55**, 223 (1933); Steacie, *et al.*, *J. Chem. Phys.*, **7**, 782 (1939); **9**, 306 829 (1941); Spinks, *Chem. Rev.*, **26**, 129 (1940).

38. Warburg and Negelein, *Z. phys. Chem.*, **102**, 236 (1922); **106**, 191 (1923); F. F. Rieke, *J. Chem. Phys.*, **7**, 238 (1939); Daniels, *et al.*, *J. Am. Chem. Soc.*, **60**, 266 (1938); **61**, 3525, 3529 (1939); Manning, *J. Phys. Chem.*, **42**, 815 (1938); Franck and Herzfeld *ibid.*, **45**, 978 (1941).
39. Norrish, *J. Chem. Soc.*, 761 (1927); Luther and Weigert, *Z. phys. Chem.*, **51**, 297; **53**, 385 (1905); Coehn, *et al.*, *Z. Elek.*, **13**, 545 (1907); *Z. phys. Chem.*, **70**, 88 (1909); **91**, 347 (1916); **105**, 356 (1923).
40. Zimmer, *ibid.*, **B23**, 239 (1933); Leighton and Lucy, *J. Chem. Phys.*, **2**, 756, 760 (1934).
41. Trautz, *Z. phys. Chem.*, **53**, 1 (1905); Bonhoeffer, *ibid.*, **116**, 391 (1925); Haber and Zisch, *Z. Phys.* **9**, 302 (1922); Polanyi, *et al.*, *ref.* (12); Emeléus, *J. Chem. Soc.*, **125**, 2491 (1924); **127**, 1362 (1925); Chariton and Walta, *Z. Phys.*, **39**, 547 (1926); Semenoff, *ibid.*, **46**, 109 (1927); Lifschitz and Kalberer, *Z. phys. Chem.*, **102**, 393 (1922); Evans, *et al.*, *J. Am. Chem. Soc.*, **45**, 278 (1923); **48**, 715 (1926); Albrecht, *Z. phys. Chem.*, **136**, 321 (1928); Kautsky, *Trans. Farad. Soc.*, **21**, 591 (1926); for general discussion of Luminescence, see *ibid.*, **35**, 1 *et seq.* (1939).

Valuable summaries on Catalysis and Photochemistry will be found in the Reports of the Committees on Catalysis and Photochemistry published in the Journal of Physical Chemistry. The Twelfth Report of the Committee on Catalysis has been published in book form.

CHAPTER XIV

SURFACE PHENOMENA

ADSORPTION

Sorption.—The surface of a liquid is in a state of strain (p. 481), or unsaturation, and that of a solid has, similarly, a residual field of force; there will consequently be a tendency for the free energy of any surface to decrease, and it is this tendency which is ultimately responsible for the phenomenon of **adsorption**. The term adsorption* refers strictly to the existence of a higher concentration of any particular component at the *surface* of a liquid or solid phase than is present in the bulk; it should be, theoretically, clearly distinguished from **absorption**, as applied to solids, since the latter refers to a more or less uniform penetration. In actual fact it is practically impossible to separate the effects of adsorption from those of absorption, particularly for systems of gases and solids, and hence the noncommittal term **sorption** is frequently employed (J. W. McBain, 1909). The word "occlusion," proposed by T. Graham (1866), in connection with his studies on the taking up of hydrogen by palladium, has a similar significance to sorption, but the former term is now generally restricted to the sorption of gases by metals.

Adsorption of Gases.—The capacity possessed by wood charcoal for the sorption of gases was observed by C. W. Scheele (1773), by the Abbé F. Fontana (1777), and others, and the earliest systematic experiments were carried out by T. de Saussure (1814) who employed a number of porous materials in conjunction with most of the gases known at the time. It is very probable that adsorption occurs on all surfaces, e.g., glass is almost invariably covered by a layer of water molecules, but the effects are not evident unless the adsorbing material, i.e., the adsorbent, is porous and has a very large area for a given mass; for this reason, various forms of charcoal have been extensively used in adsorption studies. In recent years silica gel has also been employed for experimental work as well as in industrial practice. The materials obtained by charring wood or coconut shells are able to take up large volumes of certain gases, and the amounts can be increased by subjecting the charcoal to a process of "activation"; this generally involves heating in a vacuum or an inert gas, or, better, in steam, air, chlorine or carbon dioxide, at temperatures varying from 350° to 1000° c. As a result of this treatment it appears

* The term "adsorption" was first used by H. Kayser (1881) at the suggestion of E. du Bois-Reymond.

that hydrocarbons and other impurities are removed from the charcoal, and at the same time the surface area may be increased by chemical reaction with the oxidizing agent used as activator. There is also a possibility that the degree of unsaturation of the surface atoms, in the sense considered on p. 1148, may be increased.

Experimental.—Several methods have been used to investigate the extent of the adsorption of gases under different conditions. For example, a known amount of the adsorbent is introduced into the apparatus containing a given gas, and the pressure is measured initially and when sorption is complete. The true adsorption is believed to occur rapidly and so the change of pressure within a short time is noted; the subsequent variation is assumed to be due to secondary causes, e.g., absorption, and so is ignored. If the volume of the apparatus not occupied by the adsorbent is known, the quantity of gas adsorbed can be calculated from the gas law equation. Since this volume is not always known precisely, an alternative, more direct, method has been devised (J. W. McBain, 1926–29); a small vessel containing the adsorbent is attached to a spiral spring of fused silica, or to one arm of a magnetically adjustable microbalance, and sealed up in contact with the substance to be adsorbed. The extent of sorption is determined directly from the increase of weight as indicated by the extension of the spring, or from the magnetic pull necessary to restore the balance to its original position. These procedures are of special value for the study of vapors and of gases under high pressures.

Influence of Temperature and Pressure.—Increase of pressure and decrease of temperature increase the extent of adsorption; the curves in Fig. 254, for example, show the quantities of nitrogen adsorbed by 1 g. of charcoal at various pressures and temperatures. It will be observed that, particularly at low temperatures, the adsorption of gas increases very rapidly as the pressure is increased from very small values. The considerable adsorption at low temperatures, even at small pressures, is utilized in the production of high vacua. The partly evacuated apparatus is connected to a vessel containing charcoal cooled in liquid air; as a result of the high adsorptive capacity of the charcoal under these conditions there is a considerable reduction of pressure. The decrease of adsorption with increasing temperature implies, by Le Chatelier's principle, that heat is evolved in the process of adsorption. As in the case of heats of solution, two different heats of adsorption, viz., differential and integral, must be distinguished; if accurate values were obtainable the results would probably throw much light on the mechanism of adsorption phenomena. Many attempts have been made to determine heats of adsorption experimentally, but owing to the difficulties involved reliable measurements have become available only in recent years. It should be possible, theoretically, to calculate, by means of an equation of the van't Hoff type, the differential heat of adsorption from measurements of the pressures requisite to produce a given amount of adsorption at two different temperatures. The curves showing the variation of these pressures (p) with temperature (T) are called **isosteres** and the isostere for $\log p$

against $1/T$ is linear, as it should be if the van't Hoff equation is applicable. In view of the uncertainties involved in adsorption studies, however, the heat of adsorption calculated from the slope is only approximate.

The variation of adsorption with pressure can often be represented, over a limited range of pressure, at constant temperature, by the empirical equation *

$$\frac{x}{m} = kp^{1/n}, \quad (1)$$

where x is the mass of gas adsorbed by m g. of adsorbing material at the pressure p , and k and n are constants for the given system and tempera-

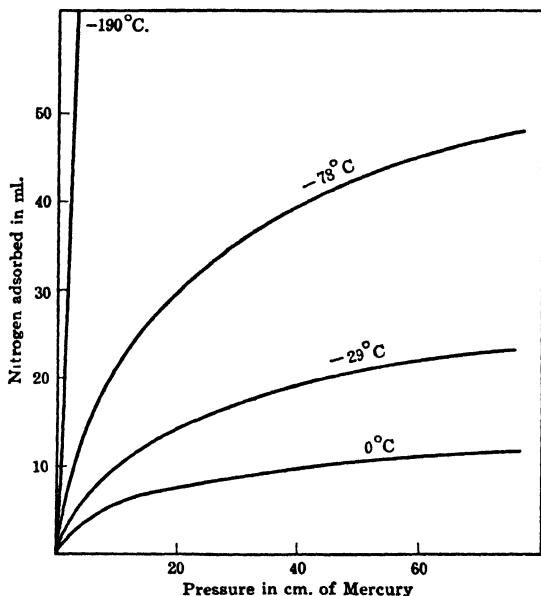


FIG. 254. Adsorption of nitrogen by charcoal

ture; this relationship will be referred to as the "classical adsorption isotherm." The quantity $1/n$ is generally less than unity, so that the amount of adsorbed gas increases less rapidly than the pressure; if it were unity, however, the adsorption equation would be equivalent to the distribution law (p. 735). Taking logarithms, (1) becomes

$$\log \frac{x}{m} = \log k + \frac{1}{n} \log p, \quad (2)$$

so that the plot of $\log x/m$, where x/m is the amount adsorbed per g. of

* The relationship is widely known as the "Freundlich adsorption isotherm"; the equation was used by earlier workers, and in fact it was originally rejected by H. Freundlich (1907) in preference to that of K. Kroeker (1892).

material, against $\log p$ should be a straight line. The actual plots show a slight curvature, especially at low temperatures, and the equations (1) and (2), although very simple and convenient, can be regarded as approximately applicable for a small range of pressure only. In view of its empirical nature, the chief use of the classical adsorption equation is as an interpolation formula, although it is still frequently employed in connection with adsorption from solution. Another isothermal equation, with a theoretical basis, will be considered shortly (p. 1198).

Nature of Adsorbent and Adsorbed Gas.—Apart from behavior of a specific type, sometimes observed, the adsorptive capacities of different substances for a given gas appear to depend primarily on their effective areas; it is found, for example, that the order of increasing extent of adsorption for a series of gases is the same for different adsorbents. There are, however, a number of exceptions attributed to the intervention of factors of a chemical nature. When these do not arise, as in the type of adsorption involving physical forces only (p. 1201), the quantities of different gases sorbed by a given solid are related in a general way to their physical properties; this is shown by the data in Table 195, where the

TABLE 195. ADSORPTION OF GASES BY CHARCOAL

Gas	Volume sorbed	Crit. Temp.	Gas	Volume sorbed	Crit. Temp.
Sulfur dioxide	380 ml.	430° K.	Carbon dioxide	48 ml.	304° K.
Chlorine	235	417°	Methane	16.2	190°
Ammonia	181	406°	Carbon monoxide	9.3	134°
Hydrogen sulfide	99	373°	Oxygen	8.2	154°
Hydrogen chloride	72	324°	Nitrogen	8.0	126°
Nitrous oxide	54	310°	Hydrogen	4.7	33°

volumes of different gases sorbed by 1 g. of active charcoal at 15° are given, together with the critical temperatures of the gases (J. W. McBain). The results show a distinct parallelism between the extent of adsorption of a gas and its critical temperature; since the latter is related to the boiling point (p. 456) and the solubility (p. 702), it can be stated, as a general rule, that the most easily liquefiable and highly soluble gases are the most readily adsorbed. The physical properties mentioned are determined essentially by the attractive forces between the molecules, and hence it is reasonable to suppose that the van der Waals or "dispersion" forces (p. 299) play an important part in adsorption.

The general conclusions described above are more particularly applicable at moderately low temperatures and at pressures that are not too near the saturation vapor pressure. As the temperature is increased there is a possibility of a different type of adsorption, involving chemical forces, which is specific in nature (p. 1201). Further, at low temperatures and at pressures approaching the vapor pressure, actual condensation may take place. The vapor pressure of a liquid in a capillary tube is lower than for a plane surface (cf. p. 496); hence condensation in small

pores can occur at pressures appreciably below the normal saturation vapor pressure for the given temperature. It should be emphasized that the temperature must be below the critical value, of course, and the pressure should be an appreciable fraction of the saturation amount.¹

Unimolecular Layers.—Until 1916 theories of adsorption postulated the existence of either a condensed liquid film or a compressed gaseous layer on the surface of the adsorbent; the films were believed to be relatively thick, becoming less dense as the distance from the surface increased, until the density was the same as in the bulk of the gas. It was pointed out, however, by I. Langmuir (1916) that because of the rapid falling off of intermolecular forces with distance, it is probable that adsorbed layers are no more than a single molecule in thickness. This view is now widely accepted for adsorption at low pressures or at moderately high temperatures. However, the adsorbed molecules can hold other gas molecules by van der Waals forces, so that multimolecular layers are possible; nevertheless, such behavior is apparent only at relatively low temperatures and at pressures approaching the saturation value (cf. p. 1202). Direct experimental evidence, based on measurements with smooth surfaces in particular, e.g., of glass, mica, platinum and mercury, has established beyond reasonable doubt that the layer of adsorbed gas is only one molecule thick at low pressures. At higher pressures, especially at temperatures below the critical point, further layers are formed, but for the purpose of the derivation to be given in the next section it will be supposed that the conditions are such that a unimolecular adsorbed layer only is present on the surface of the solid.

Langmuir's Adsorption Isotherm.—As indicated on p. 1118, Langmuir treated adsorption in a unimolecular layer from the standpoint of the equilibrium between gas molecules striking the surface, and those which evaporate off after the lapse of a certain time. If μ is the number of molecules of gas striking 1 sq. cm. of surface per sec., and α is the (constant) proportion which adheres, then $\alpha\mu$ molecules adhere to each sq. cm. of surface per sec. If θ is the fraction of the total available surface covered with gas molecules at any instant, then $1 - \theta$ is the fraction which is bare, and $(1 - \theta)\alpha\mu$ is the actual rate of condensation * of gas molecules per sq. cm. of total surface. The rate of evaporation of molecules will be proportional to the area covered, and hence is $\nu\theta$, where ν is a constant for the given gas and surface. At adsorption equilibrium the rates of evaporation and condensation of molecules will be equal; hence,

$$(1 - \theta)\alpha\mu = \nu\theta; \quad (3)$$

$$\therefore \theta = \frac{\alpha\mu}{\nu + \alpha\mu}. \quad (4)$$

* The word condensation is here used, as on p. 1118, in a general sense, and does not, of course, imply liquefaction.

If it is assumed that a uniform layer of gas only one molecule thick can form on the surface of the adsorbent, the fraction θ is proportional to the amount of gas x adsorbed by a definite mass m of adsorbent. Further, μ is proportional to the gas pressure (*p*. 278), and so (4) may be written

$$\frac{x}{m} = \frac{k_1 k_2 p}{1 + k_1 p}, \quad (5)$$

where k_1 and k_2 are constants for the given system and p is the gas pressure; (5) is known as **Langmuir's adsorption isotherm**. It may be written in the form

$$\frac{p}{x/m} = \frac{1}{k_1 k_2} + \frac{p}{k_2} \quad (6)$$

so that if $p/(x/m)$ is plotted against p a straight line should be obtained; the data represented in Fig. 255 (J. W. McBain and G. T. Britton, 1930)

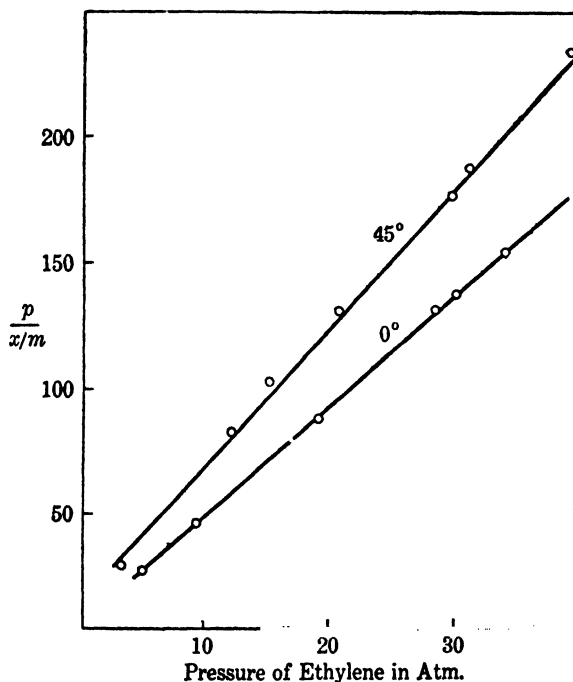


FIG 255. Langmuir adsorption isotherms

show satisfactory agreement with this requirement. Deviations have been observed which may be attributed to various causes, e.g., non-uniformity leading to different types of adsorption on the same surface.

At very low pressures, or for a poor adsorbent, only a small fraction of the surface is covered with molecules; θ is then very small and $1 - \theta$ is almost equal to unity, so that (3) reduces to

$$\nu\theta = \alpha\mu \quad (7)$$

or

$$\frac{x}{m} = k_1 k_2 p.$$

Under these conditions the amount of gas adsorbed is directly proportional to the pressure, as has in fact been found in certain cases for very low gas pressures (cf. Fig. 254). On the other hand, when the adsorption is considerable, so that θ is almost unity, (3) may be written

$$(1 - \theta)\alpha\mu = \nu; \quad (8)$$

$$\therefore \theta = 1 - \frac{\nu}{\alpha\mu}$$

and

$$\frac{x}{m} = k_2 - \frac{k_2}{k_1 p}. \quad (9)$$

As the pressure p is increased $k_2/k_1 p$ becomes smaller and the amount of gas adsorbed, x/m , should tend toward a limiting value; this corresponds to the state when the surface is virtually completely covered with a single layer of gas molecules.

The Langmuir isotherm reduces to the form $x/m = kp$ at low pressures and to $x/m = k$ at relatively high pressures; hence at intermediate pressures an expression of the type

$$\frac{x}{m} = kp^{1/n},$$

where $1/n$ lies between 0 and 1, may be expected to hold. This is in fact the classical adsorption isotherm, which is thus a special case of the Langmuir equation, applicable over a limited range of pressure.

It is important to point out that the agreement between the Langmuir isotherm and the results of experiment cannot be taken as definite proof of the theory on the basis of which (5) was deduced, for the same relationship can be derived in other ways. Nevertheless, the mechanism of adsorption described above has many advantages, particularly as regards the simplicity with which it can be applied to surface reactions (Chapter XIII), and related phenomena.²

Types of Adsorption.—A study of absorption phenomena has established the fact that there are two main categories of adsorption, depending on whether the association between the gas and the solid surface is physical or chemical in character; that is to say, according as van der Waals (or dispersion) forces or attractions analogous to valence, respec-

tively, are involved. The first type, known as **van der Waals adsorption**, is characterized by relatively small heats of adsorption, viz., about 5 kcal. per mole, or less, which are of the same order as heats of vaporization. The equilibrium between solid and gas is reversible, and is rapidly attained when the temperature and pressure are changed. If the rate of adsorption of the gas can be regarded as a process requiring an energy of activation (see Chapter XIII), then the rapid establishment of equilibrium at low temperatures means that the activation energy is small, probably of the order of 1 kcal., or less, per mole.

All gases exhibit van der Waals adsorption, but in some cases the second type of adsorption, called **chemisorption** since it involves forces of a chemical nature, may ensue, especially at higher temperatures. The fact that the forces are similar to those concerned in chemical combination is indicated by the heats of adsorption which are of the order of 20 to 100 kcal. per mole. The bonds formed between the material of the surface and the adsorbed gas are thus almost as strong as those existing in stable stoichiometric compounds. In some instances the rate of chemisorption is considerable even at low temperatures, e.g., hydrogen on tungsten, so that the energy of activation for the adsorption process is very small, but in others activation energies of 5 to 20 kcal., or more, are required. In cases of the latter type, the process has been referred to as **activated adsorption** (H. S. Taylor, 1931).

If activated adsorption is possible, for a given gas and surface, the transition from van der Waals adsorption to chemisorption may sometimes be observed as the temperature is raised. For example, the amount of hydrogen adsorbed on a nickel surface is large at -200°C .; it decreases sharply as the temperature is raised, but increases again toward a maximum value, at about -100° , and subsequently falls off slowly. Somewhat similar, but more complicated, results have been obtained in connection with the adsorption of hydrogen on a zinc oxide surface. In each case there is evidence that the rate of adsorption at very low temperatures is greater than at higher temperatures.

Since the extent of any particular kind of adsorption must decrease as the temperature is raised, adsorption being accompanied by the evolution of heat, it is evident that the high temperature adsorption must be essentially different from that occurring at low temperatures. The former is mainly activated chemisorption, while the latter is van der Waals adsorption. Physical forces are nonspecific in nature; hence, van der Waals adsorption occurs in all cases, but chemisorption will only be observed in addition when some kind of chemical interaction is possible. If the latter process requires an appreciable energy of activation, it will become apparent only as the temperature is raised, and the molecules held by van der Waals forces acquire the necessary energy. This is evidently the case in the two instances referred to above, when the total amount of adsorption is found first to decrease and then to increase, to some extent, as the temperature is raised. It should be noted that theoretically ac-

tivated adsorption should occur even at low temperatures, but the rate may be so small that the adsorption is not apparent.

Van der Waals Adsorption.—At relatively low pressures, and particularly at moderately high temperatures, it is probable that molecules adsorbed by van der Waals forces consist of a single molecular layer. Adsorption isotherms of the classical or Langmuir types are then satisfactory to account for the variation of adsorption with pressure. At low temperatures, and more especially as the pressures approach the saturation vapor pressure of the liquid at the experimental temperature, various types of behavior have been observed with different systems. Five such types, of which examples are known, are depicted in Fig. 256; the broken

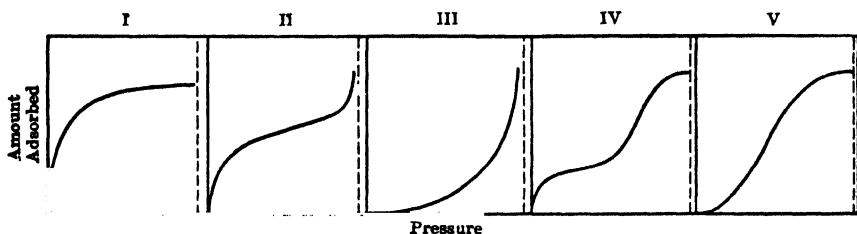


FIG. 256. Types of van der Waals adsorption

(vertical) line indicates the saturation vapor pressure. The isotherm I is of the Langmuir form, and represents the behavior of nitrogen (at -195°C.) or of oxygen (at -183°C.) on charcoal; II is exhibited by nitrogen on iron catalysts, on silical gel, and other surfaces at -195°C. ; III is the type for bromine (at 79°C.) and iodine on silica gel; IV is the form for benzene on ferric oxide gel at 50°C. ; and V is for water vapor on charcoal at 100°C. These results have been interpreted by S. Brunauer, *et al.*, (1938-39) on the basis of the formation of a single adsorbed layer at low pressures followed by the attachment of further layers as the pressure is increased. By applying the Langmuir treatment in a generalized form to the concept of the adsorption of multimolecular layers, the expression

$$v = \frac{v_m c f}{1 - f} \cdot \frac{1}{1 + (c - 1)f} \quad (10)$$

has been derived for adsorption on a free surface, without restriction as to the number of layers; v is the volume of gas adsorbed at a given pressure p (and constant temperature), f is equal to p/p^0 , where p^0 is the vapor pressure at the experimental temperature, v_m is the volume of gas adsorbed when the surface is completely covered by a unimolecular layer, and c is approximately equal to $e^{(E_1 - E_L)/RT}$, where E_1 is the heat of adsorption of the gas in the first layer and E_L is the heat of liquefaction of the gas. If E_1 is greater than E_L , so that c is considerably greater than

unity, (10) gives a curve of type II, but if E_1 is less than E_L it is of the form of III in Fig. 256. In the special case of a single unimolecular layer, (10) becomes equivalent to the Langmuir equation (6), and hence represents curve I. It will be seen that IV and V differ from II and III, respectively, in so far as the adsorption, in the former cases, reaches a limit at pressures well below the vapor pressure. By extending the multimolecular layer theory so as to allow for the possibility that the capillary pores of the surface become filled, possibly as a result of condensation, at pressures appreciably below the saturation value, types IV and V have been explained. The condition for IV, as for II, is that $E_1 > E_L$, while for V it is the same as for III, i.e., $E_1 < E_L$.

By rearrangement, (10) takes the form

$$\frac{f}{v(1-f)} = \frac{1}{v_m c} + \frac{(c-1)}{v_m c} f$$

so that the plot of $f/v(1-f)$ against f should be a straight line; the slope will be equal to $(c-1)/v_m c$, and the intercept on the ordinate axis will be $1/v_m c$. From these two quantities the value of v_m , that is, the volume of gas required to give a complete unimolecular layer, can be determined. Since the number of molecules of gas in this volume is known, and the area of cross section per single molecule is available (Chapter IV), it is possible to calculate the area of the adsorbing surface. The same information can be obtained in case I directly from the limiting adsorption, and in cases II and IV by extrapolating the intermediate (linear), flattish portion of the isotherm to zero pressure; the corresponding volume of gas, approximately equivalent to the point at which the isotherm first commences to flatten, represents that required to form a complete layer one molecule in thickness. The foregoing procedures, especially using nitrogen as the adsorbed gas, have been used to determine the surface areas of catalysts and other adsorbing materials (S. Brunauer and P. H. Emmett, 1935 *et seq.*).

Chemisorption: Activated Adsorption.—As mentioned above, the type of adsorption known as chemisorption involves forces of a chemical nature. Some evidence for this view is provided by the fact that the adsorption is often irreversible. One of the best examples of chemical association between the adsorbing material and the adsorbed substance is provided by the work of I. Langmuir (1916) on the adsorption of oxygen on tungsten; a very stable film is formed on the tungsten surface that is not the normal oxide for it has different chemical properties. In the course of time, however, the oxide WO_3 distils from the tungsten surface and is deposited on the walls of the vessel. The adsorbed oxygen film is presumably held to the tungsten by chemical forces, but since the atoms of the latter are not completely severed from those in the body of the metal a "surface compound" results; when the tungsten detaches itself from the other atoms in its proximity, the normal compound WO_3 can

form. Similar behavior has been observed in the adsorption of carbon monoxide on tungsten, when a carbonyl is eventually produced, and of oxygen on carbon, when the monoxide and dioxide are obtained by raising the temperature and pumping off the adsorbed gas. Chemisorption also occurs with hydrogen and carbon monoxide on zinc oxide-chromium sesquioxide mixtures, hydrogen on nickel and other surfaces, and oxygen on silver, gold, platinum and carbon; in the latter instances definite stoichiometric chemical compounds are formed.

Activated adsorption, e.g., hydrogen on nickel, zinc oxide, etc., is usually observed at high temperatures when the extent of van der Waals adsorption is small; hence a single molecular layer only is formed and, apart from disturbing factors, the Langmuir isotherm is obeyed. Although the word "activated" was not used to imply any particular activity of the adsorbed substance, yet when hydrogen undergoes activated adsorption it is apparently dissociated into atoms on the surface. Substances, e.g., nickel, zinc oxide, etc., on which activated adsorption of hydrogen occurs are good catalysts for the para-ortho hydrogen conversion, and for hydrogenation reactions in general which probably involve hydrogen atoms. It is likely that chemisorption is involved in all cases of surface catalysis, especially when the activation energy is markedly less than for the corresponding homogeneous gas reaction (cf. p. 1146).³

Persorption.—Reference may be made here to sorption of an entirely different character. It is known from X-ray examination that the water molecules in zeolites are not held by ordinary valence bonds, but merely fit into the vacant spaces in the lattice of aluminum, silicon, oxygen and metal atoms (p. 396). On dehydration the water is removed and the spaces can be filled by other molecules; dehydrated zeolites are, therefore, good sorbing agents and the properties of *chabazite*, $\text{CaAl}_2\text{Si}_4\text{O}_{12} \cdot 6\text{H}_2\text{O}$, have been particularly investigated in this connection. Partially dehydrated *chabazite* readily takes up vapors of water, and methyl and ethyl alcohols, although benzene, ether and acetone vapors are not sorbed (O. Weigel and E. Steinhoff, 1925). The difference in behavior of the substances in these two groups is attributed to their dimensions, for the *chabazite* acts as a true molecular sieve permitting only the entry of molecules smaller than the vacant spaces in the lattice originally occupied by the water molecules. From an examination of the sorption of a number of substances by *chabazite* the diameter of the pores has been estimated as 3.5 Å. (O. Schmidt, 1928). To distinguish between this and other types of adsorption J. W. McBain (1930) has suggested the term **persorption**; it differs from true solid solution only in the respect that it is not completely uniform throughout the solid. It is probable that persorption is operative in the highly active charcoals, and perhaps to some extent in silica gel; the occlusion of hydrogen by palladium (p. 784) may well be an extreme case of combined activated adsorption and persorption. It may be mentioned that the heat evolved when carbon dioxide is adsorbed by dehydrated zeolites is nearly 13 kcal., compared with 7 to 8 kcal. for adsorption on charcoal and silica gel (A. B. Lamb, 1935).

Heat of Adsorption.—Measurements of heats of adsorption have been made over a period of years, but the experimental difficulties were not fully

realized until recently. Many of the results reported are inaccurate because of the low heat conductivity of the adsorbent and the consequent time lag in the transfer of heat to the temperature measuring device, and also because adsorption was not uniform throughout the material employed. Modern workers have designed apparatus with the object of overcoming these sources of error; the gas which is being adsorbed is led into the center of the adsorbent and special precautions are taken to ensure good thermal contact. Two main types of calorimeter have been employed for this work: first, a form of the Bunsen ice calorimeter, in which the heat produced is determined by the volume change accompanying the melting of ice, and second, special forms of apparatus in which the rise of temperature is measured by a thermocouple or resistance thermometer. Small amounts of the experimental gas are allowed to enter the adsorption chamber and the heat evolved is determined; the amount of gas adsorbed is estimated from the change of pressure in the apparatus. This procedure is repeated several times, and from the results the approximate differential heats of adsorption at various stages can be calculated per mole of gas adsorbed. In spite of many improvements in technique the experimental values can still not be regarded as highly accurate; definite tendencies have, nevertheless, been clearly established.

It had been suggested at one time that the differential heat of adsorption increased to a maximum and then diminished with increasing extent of adsorption. Experimental errors were, however, responsible for this conclusion and it is now known that heats of adsorption almost invariably decrease steadily right from the commencement of adsorption. The reason for this is not clear: it may be due to the fact that the first quantities of gas to be adsorbed will attach themselves to the most active parts of the surface, or that more than one type of adsorption is occurring on the surface. When oxygen is adsorbed on carbon, and possibly also on platinum, the initial heat evolution may be 80 kcal. per mole, or even higher; with increasing adsorption, however, the value falls off rapidly to about 20 kcal. There is little doubt that in this instance the adsorption involves chemical combination between the gas and the adsorbent.⁴

Adsorption at Surfaces of Solutions.—As a result of the tendency of the free energy of a surface to decrease, the concentration of a solute on the surface may differ from that in the bulk of the solution. The component with the lower surface tension will tend to concentrate in the surface, for in this manner the free (surface) energy of the system is reduced. The complete replacement of solvent by solute molecules in the surface will, of course, be prevented by thermal agitation and forces of molecular attraction. It follows, therefore, from general considerations, that if a particular solute lowers the tension at a given interface, there will be a greater proportion of solute to solvent at the interface than in the bulk of the solution. This will consequently represent a case of adsorption of the solute at the surface of the solution. Substances which produce a marked reduction in interfacial tension are said to be **surface active**. On the other hand, if the solute brings about an increase of surface tension, its concentration in the surface will be *less* than in the remainder of the solution; this type of behavior, called **negative adsorption**, is shown particularly by electrolytes. Since a substance which

lowers surface tension accumulates at the interface, it is evident that the surface tension of the solution will tend to approach that of the solute; on the other hand, when the surface tension is raised, the solute is rejected from the surface and so the surface tension of the solution will differ little from that of the pure solvent (cf. p. 496).

The Gibbs Adsorption Equation.—The exact relationship between adsorption and surface tension was first derived by J. Willard Gibbs (1878), and independently by J. J. Thomson (1888); it is known as the **Gibbs adsorption equation**. Many deductions of varying rigidity and complexity are to be found in the literature; the following is based on the method given by Gibbs, involving the use of the thermodynamic potential.

When considering free energy changes in Chapter III no account was taken of the possibility of changes in the surface energy; in fact, the tacit assumption involved was that there was no alteration in the surface area when the system underwent a change. In order to allow for the possibility of a change in free energy resulting from an increase or decrease of the exposed surface, it is necessary to include a term γs , where γ is the interfacial tension, i.e., interfacial energy per sq. cm. (cf. p. 481), and s is the surface area. Apart from the surface contribution, the free energy of a system of two components is given by (155), p. 237, as $F = \mu_1 n_1 + \mu_2 n_2$, where n_1 and n_2 are the respective numbers of moles. Addition of the γs term yields

$$F = \gamma s + \mu_1 n_1 + \mu_2 n_2,$$

and upon differentiation it follows that

$$dF = \gamma ds + s d\gamma + \mu_1 dn_1 + n_1 d\mu_1 + \mu_2 dn_2 + n_2 d\mu_2. \quad (11)$$

Alternatively, dF is obtained by adding the term γds , for an increase of ds in the surface area, to (165), p. 238, so that

$$dF = -SdT + VdP + \gamma ds + \mu_1 dn_1 + \mu_2 dn_2, \quad (12)$$

and by comparing (11) and (12), the result is

$$SdT - VdP + s d\gamma + n_1 d\mu_1 + n_2 d\mu_2 = 0, \quad (13)$$

or at constant temperature and pressure,

$$s d\gamma + n_1 d\mu_1 + n_2 d\mu_2 = 0. \quad (14)$$

Imagine the system under consideration to be divided into two parts, one consisting of all that portion which comes under the influence of surface forces, and the other the remainder of the solution. The former may be called the "surface phase," and the latter the "bulk phase," which is quite free from surface effects. If n_1^0 and n_2^0 are the numbers of moles of the two components in the bulk phase, corresponding to n_1 and n_2 in the surface phase, then while (14) holds for the latter, the relationship

$$n_1^0 d\mu_1 + n_2^0 d\mu_2 = 0 \quad (15)$$

applies to the bulk phase; this is, of course, a form of the Gibbs-Duhem equation, applicable when surface considerations are not involved. If (15) is multiplied by n_1/n_1^0 and subtracted from (14), it is seen that

$$sd\gamma + (n_2 - n_1 n_2^0/n_1^0)d\mu_2 = 0; \quad (16)$$

$$\therefore -\frac{d\gamma}{d\mu_2} = \frac{n_2 - n_1 n_2^0/n_1^0}{s}. \quad (17)$$

The quantity n_2 is the number of moles of one component, e.g., the solute, associated with n_1 moles of solvent in the surface phase, and $n_1 n_2^0/n_1^0$ is the corresponding number of moles associated with n_1 moles of solvent in the bulk phase; the right-hand side of (17) may, therefore, be regarded as the excess concentration of solute per unit area of surface. This excess concentration is given the symbol Γ_2 and is called the surface concentration of solute per unit area of interface. It is not strictly a concentration term, since it is a number of moles divided by an area, but it is nevertheless a definite quantity defined by the right-hand side of (17). It is important to point out that in spite of the use of the arbitrary amount n_1 to define Γ_2 , it is independent of this quantity, since the left-hand side of (17) depends only on the nature of the surface phase and not on its amount. The actual amount of the surface phase, therefore, does not affect the value of Γ_2 , provided that at least the whole of the portion of the system coming under the influence of surface forces is included. Introducing the symbol Γ_2 in (17), this becomes

$$\Gamma_2 = -\frac{d\gamma}{d\mu_2}. \quad (18)$$

Since $\mu_2 = \mu_2^0 + RT \ln a_2$ (p. 684), where a_2 is the activity of the solute, then at constant temperature $d\mu_2 = RT \ln a_2$, and substitution in (18) gives

$$\Gamma_2 = -\frac{1}{RT} \cdot \frac{d\gamma}{d \ln a_2} = -\frac{a_2}{RT} \cdot \frac{d\gamma}{da_2}. \quad (19)$$

These are two forms of the Gibbs equation. Although in practice it is usually applied to the solute, it can be seen from the deduction that it should hold equally for either component of a binary system; the subscripts can, therefore, be neglected, giving

$$\Gamma = -\frac{a}{RT} \cdot \frac{d\gamma}{da}. \quad (20)$$

In deriving (20) no assumption has been made as to the nature of the system or of the surface; its most obvious application is, however, to a liquid solution, the surface being that between the liquid and its vapor, or possibly air. It might also be reasonably assumed to apply to an interface between the solution and a liquid, e.g., mercury, with which it is

completely immiscible; the quantity γ then refers to the interfacial tension. The treatment given above can be extended, if required, to a system involving partially miscible liquids, when the same general result is obtained. Assuming, for the present, that the activity of the solute is proportional to its concentration, it follows from (20) that if $d\gamma/da$, i.e., the change of interfacial tension with increase of concentration, is negative, then Γ is positive; the concentration of the solute will therefore be greater in the surface phase than in the bulk phase, as deduced above on general grounds. For dilute solutions, the activity may be replaced by concentration c , and the approximate form

$$\Gamma = - \frac{c}{RT} \cdot \frac{d\gamma}{dc} \quad (21)$$

of the Gibbs equation obtained in this way is frequently employed.

Verification of the Gibbs Equation.—The direct experimental verification of the Gibbs equation is very difficult. The first attempts (W. C. McC. Lewis, 1908–09) were made with aqueous solutions of substances which lower interfacial tension, e.g., sodium glycocholate, caffeine and various dyestuffs, and mercury or a hydrocarbon oil. A stream of drops of oil was passed through the aqueous solution and the bulk concentration determined at the beginning and end; the difference gave the amount of solute adsorbed at the surface of the drops. If the dimensions of the latter and the total quantity of oil are known the total area can be calculated and hence the concentration per unit area may be evaluated. The observed concentration changes were very small and the results not very accurate, but on the whole it appeared that the amount adsorbed was somewhat greater than that calculated from the Gibbs equation on the assumption that concentrations could be used instead of activities. Similar results were obtained by F. G. Donnan and J. T. Barker (1911) who studied the adsorption of nonylic acid and saponin from aqueous solution at the liquid-air interface. Bubbles of air, of known size and number, and hence of known total area, were passed through the solution, and the concentration of the liquid obtained when the bubbles collapsed was taken as that of the surface phase. This experimental method was refined by J. W. McBain, *et al.* (1927–29), and observations were made on aqueous solutions of camphor, amyl alcohol and *p*-toluidine, but the extent of adsorption was found to be from two to four times that to be expected from the Gibbs equation. Further investigation has shown that although reproducible results can be obtained by the gas bubble method under a given set of conditions, alteration of the size and speed of the bubbles yields different adsorption values. The source of the discrepancy is that moving surfaces are involved, whereas the Gibbs equation can hold only for a static surface. In the procedure used by J. W. McBain and R. C. Swain (1936) to eliminate this error, the ends of a trough containing the solution were coated with the hydrocarbon dicetyl, so that the

liquid bulged about 0.05 to 0.1 mm. above the upper level of the vessel without overflowing. By means of a microtome blade which was moved rapidly across the liquid the surface layer was cut off and collected in a silver-lined cylinder attached to the blade. The difference in concentration between the bulk of the liquid and that constituting the surface was determined by an interferometer. The results for solutions of phenol and hydrocinnamic acid, which show positive adsorption, and for sodium chloride, which is negatively adsorbed at the air-liquid interface, were in excellent agreement with the requirements of the Gibbs equation. Another method of testing this equation involves a sudden compression of the surface into the bulk of the liquid in the path of an interferometer beam. By this means the change in concentration, due to the extra amount in the surface, can be determined, and hence the composition of the surface may be calculated (J. W. McBain, 1940). These procedures provided the first direct verification of the Gibbs equation; an indirect method is given below (p. 1213).⁵

Molecular Orientation in the Surface.—The variation of the surface tensions of relatively concentrated aqueous solutions of soluble fatty acids with concentration can be represented by an equation of the form proposed by B. Szyszkowski (1908), viz.,

$$\frac{\gamma}{\gamma_0} = 1 - X \ln \frac{c}{Y}, \quad (22)$$

where γ and γ_0 are the surface tensions of solution of concentration c and of pure water, respectively, and X and Y are constants; X is a constant for a series of fatty acids containing from two to six carbon atoms, but Y decreases with increasing length of the hydrocarbon chain. If (22) is differentiated with respect to $\ln c$, it follows that

$$\frac{d\gamma}{d \ln c} = -X\gamma_0 \quad (23)$$

and if this value of $d\gamma/d \ln c$ is substituted in the approximate form of the Gibbs equation, in which concentrations replace activities, it is seen that

$$\Gamma = \frac{X}{RT} \gamma_0. \quad (24)$$

Since X is the same for a series of fatty acids and γ_0 is constant, it is evident that in the fairly concentrated solutions considered, the excess of fatty acid in the air-solution interface becomes constant and independent of the nature of the acid. Such a result is best explained by supposing that the material adsorbed at the interface forms a single layer of molecules, and that as the concentration of fatty acid in the bulk phase increases the amount adsorbed increases until a complete unimolecular layer is formed. If the molecules are oriented vertically, presumably with

the carboxyl groups pointing inward toward the bulk phase, and the hydrocarbon chains outward, the amount adsorbed, expressed in moles per sq. cm., will be independent of the length of the chain: this is in harmony with (24). If the condition for which this equation applies is assumed to represent a complete layer, the area occupied by each mole in the surface is equal to $1/\Gamma$; the area A occupied by a single molecule would then be $1/N\Gamma$, where N is the Avogadro number. From (24), therefore,

$$A = \frac{RT}{NX\gamma_0} = \frac{kT}{X\gamma_0}, \quad (25)$$

where k is the Boltzmann gas constant, 1.38×10^{-16} erg per degree. From the work of Szyszkowski the value of X is known to be 0.179, and taking the surface tension of water as 72.5 dynes per cm., the area occupied by a single molecule of fatty acid is found to be 31×10^{-16} sq. cm. This figure is slightly higher than that obtained by other methods (cf. p. 1212), but it is not surprising since no allowance was made for the molecules of water which remain in the surface layer.

The above considerations indicate the probability of the formation of unimolecular layers at the surface of a solution containing a substance producing a marked lowering of surface tension. This view applies not only to fatty acids but to other substances, particularly those having molecules which tend to orient themselves at the surface because they possess a polar group. It is possible, however, that this condition represents an ideal, and there may be some tendency, in certain cases, for thicker layers to form to a limited extent.

FILMS OF INSOLUBLE SUBSTANCES

Formation of Unimolecular Surface Films of Spreading Oils.—A film of an insoluble substance possessing a polar group, and which therefore tends to spread on the surface of water (cf. p. 485), may be regarded in a sense as an extreme case of surface adsorption. It is indeed true that there is no sharp transition from the surface layers of solutions of acetic, propionic, butyric and other soluble fatty acids, described above, to the films produced by the insoluble acids with long hydrocarbon chains. The two types of systems, however, lend themselves to different methods of study.

The discovery was made by Miss A. Pockels (1891) that the area covered by a spreading oil, e.g., oleic acid, on the surface of water could be varied at will by confining the film between movable barriers placed across a shallow tray filled with water on which the oil was placed. The observations were continued by Lord Rayleigh (1899), who concluded that the oil spread out into a film one molecule in thickness. Subsequent work by H. Devaux (1904, 1912) confirmed these results, but modern work on the subject may be said to commence with the improvements in

the experimental methods made by I. Langmuir (1917) and later by N. K. Adam (1921, 1926).⁶

A known weight of the spreading oil, preferably a pure compound, is dissolved in benzene and placed on the surface of clean water in a shallow, rectangular trough, seen in plan in Fig. 257. The benzene evaporates leaving a film of the oil which is confined between a movable barrier *A*, consisting of a waxed paper or a glass strip, placed right across the trough and a similar floating barrier *B* that remains at one end. Leakage of the oil film past the ends of *B* was prevented in Langmuir's original apparatus by air jets, but in Adam's improved form (1926) thin strips of platinum foil *C* were employed. Langmuir measured

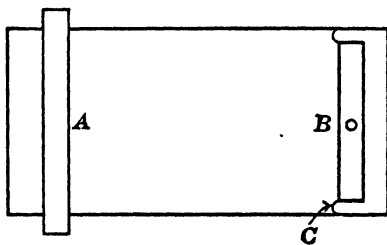


FIG. 257. Study of surface films

the force exerted by the film covering the area between *B* and *A* by attaching a lever and balance-pan (not shown) to the center of the float *B* and finding the weights required to bring *B* back to the position it occupied in the absence of the oil film. In Adam's apparatus the force was measured by a torsion wire and optical-lever system, permitting values as low as 0.01 dyne per cm. to be determined. The method of experiment was to move the barrier *A* in stages toward *B*, thus decreasing the area occupied by the film, and to measure the surface force exerted at each stage. Since the weight of the film-forming material and its molecular weight are known, the corresponding area occupied by a single molecule can be calculated, utilizing the Avogadro number.

Structure of Surface Films.—The dependence of the surface force exerted by an oil film on the average area occupied by each molecule in it is shown in Fig. 258, which is typical for a saturated fatty acid on distilled water. It is seen that for relatively large areas the surface force is very small, but at the point *a* it commences to increase, at first slowly and then, from *b* onward, extremely rapidly until at *c* the film collapses. At this point the film, which has hitherto been invisible, commences to crumple up and strain lines can be seen on the surface. If the portion *bc* of the force-area curve is extrapolated backward to zero compression of the film (*O*), the corresponding area is 20.5 sq. Å., i.e., 20.5×10^{-16} sq. cm., per molecule. It is a very remarkable fact that similar curves are obtained for a series of fatty acids, independent of their chain length, and for long-chain amides, alcohols, methyl ketones, monobasic esters, and other substances with polar groups; in each case the extrapolated minimum area per single molecule at zero compression is very close to 20.5 sq. Å., as for the fatty acids. The simplest interpretation of these results is that the oil films consist of a single layer of molecules, often called **monolayers**, arranged more or less vertically with the polar groups attached to the surface of the water, and the hydrocarbon chain pointing outward. As the film is compressed, by moving the enclosing barriers nearer together, the molecules pack more tightly, but a point is reached, approximately

at *b* in Fig. 258, when the oriented molecules are so closely packed that any further attempt at compression demands the exertion of an appreciable surface force; the value of the latter thus increases rapidly from *b*

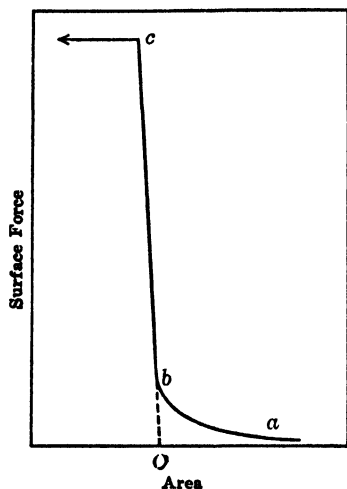


FIG. 258. Force-area curve for fatty acid on water

to *c* for a small decrease of area. At *c* the vertical molecules are pressed so tightly against each other that further compression causes them to pile up on top of one another. If this argument is correct, the limiting area occupied by each molecule will be determined by the cross-sectional area of the hydrocarbon chain, and hence will be independent of its length or of the nature of the polar end-group, provided that the latter is sufficiently attracted by the water for the chains to be "anchored" to the surface. The constant area of 20.5 sq. A., at zero compression, is in harmony with expectation, especially as this figure is in good agreement with the cross section of a hydrocarbon chain as estimated by X-ray (p. 399) and other methods. The force-area curves for fatty acids spread on the surface of dilute hydrochloric acid have two portions: the first corresponds to a limiting area of 25 sq. A. per molecule and the second to the usual value of 20.5 sq. A. It has been suggested that the former represents close packing of the carboxyl groups in the surface of the water, but with increasing compression the molecules are supposed to rearrange themselves by pushing some of the end-groups further into water, leaving the chains closely packed.

Expanded and Gaseous Films.—The films described above, consisting of closely packed single layers of oriented molecules, have been called **condensed films** to distinguish them from **expanded films** formed as the temperature is raised; the longer the hydrocarbon chain the higher the temperature at which "expansion" occurs. In these films the area occupied per molecule is greater than in the condensed films; this appears to be due to the mutual repulsion of the end-groups. Certain films, known as the **liquid expanded type**, have force-area curves very similar to the pressure-volume curves obtained for a gas undergoing liquefaction (see Fig. 91); there is a range of area over which they exhibit a definite force equivalent to a **surface vapor pressure** which is independent of the total area occupied. At low surface pressures, and especially at elevated temperatures, when the molecules are relatively far apart, **gaseous films** are sometimes formed; these bear a striking analogy to three-dimensional

If the product of the force F^* and the average area A occupied by

* This should not be confused with the free energy.

each molecule, i.e., FA , is plotted against F , curves are obtained similar to the PV - P curves for gases. In the gaseous films the molecules probably lie more or less flat on the surface; hence substances with polar groups at each end, e.g., esters of dibasic acids, form such films most readily.⁷

Relationship between Films of Soluble and Insoluble Substances.—It is evident that condensed unimolecular films, with their oriented molecules, are equivalent to an extreme case of adsorption of a soluble substance (cf. p. 1210); it is in connection with gaseous films that a quantitative agreement becomes apparent. The force F exerted by an insoluble film, as measured by the method of Langmuir or of Adam, is really the difference between the surface tensions of pure water on one side of the floating barrier and of the water covered with oil film on the other side. This definition may be extended to the surface layers of soluble substances, so that F is equal to the difference between the surface tension of the pure solvent γ_0 and of the solution γ , i.e.,

$$F = \gamma_0 - \gamma. \quad (26)$$

Differentiating with respect to c ,

$$\frac{dF}{dc} = - \frac{d\gamma}{dc}; \quad (27)$$

$$\frac{F}{c} \cdot \frac{d \ln F}{d \ln c} = - \frac{d\gamma}{dc}. \quad (28)$$

If the left-hand side is substituted for $-d\gamma/dc$ in the approximate Gibbs equation (21), then

$$\Gamma = \frac{F}{RT} \cdot \frac{d \ln F}{d \ln c}. \quad (29)$$

The area A occupied per single molecule is equal to $1/N\Gamma$, by the arguments on p. 1210, and substitution in (29) gives

$$FA = \frac{RT}{N} \cdot \frac{d \ln c}{d \ln F} = kT \frac{d \ln c}{d \ln F}, \quad (30)$$

where k is the Boltzmann constant. It is thus possible, from measurements of the surface tension of a solution at various concentrations, to plot FA against F , the F being equal to $\gamma_0 - \gamma$, and the FA terms being evaluated by (30). The curves obtained, like those for gaseous films of insoluble substances, are of exactly the same type as the PV against P curves for gases. The curves derived by the two methods, for soluble and insoluble substances respectively, show a continuous transition without a break, as the length of the hydrocarbon chain is increased. The acid with twelve carbon atoms is sufficiently soluble for F to be obtained from surface tension measurements, and at the same time it is sufficiently insoluble to give a surface film permitting the surface pressure to be determined directly; the two sets of data were found to fall on the same FA against F curve. Since the results for the solution were obtained by

(30), based on the Gibbs adsorption equation, the agreement may also be regarded as a strong, if indirect, proof of the accuracy of the latter.*

Surface Layers and Gas Laws.—The resemblance between gaseous films of both soluble and insoluble substances and ordinary three-dimensional gases is more than a qualitative one. If the molecules in the surface film are supposed to be free to move in any direction, within the two dimensions of the surface, and it is assumed, further, that the actual cross-sectional area of the molecules is small in comparison with the area they inhabit and that they do not attract each other, then it can be readily deduced by the methods of the kinetic theory of gases * that

$$FA = kT, \quad (31)$$

where A is the average area occupied by a single molecule and k is the Boltzmann constant. This equation is the quantitative equivalent of the gas law equation $PV = kT$, where V is the average volume inhabited by a *single* molecule. As with gases, (31) is obeyed at low values of the surface pressure F , equivalent to low gas pressures; under these conditions the limiting values of FA are in excellent agreement with the requirements of (31), viz., at 17°C , kT is $1.38 \times 10^{-16} \times 290 \approx 400 \times 10^{-16}$ erg, for surface layers of both soluble and insoluble substances. At higher surface pressures an equation similar to the one proposed by Amagat for gases may be employed; thus,

$$F(A - b) = kTx, \quad (32)$$

where b allows for the cross-sectional area of the molecules and x for their mutual attraction; as the attraction increases x decreases. For a series of soluble acids x decreases with increasing chain length, suggesting an increase of mutual attraction between the chains.

A relationship similar to (32) holds for a unimolecular film of benzophenone on mercury, the surface force F being determined from the difference in surface tension of mercury with and without the surface film; owing to the small attraction between the molecules in this instance the value of x may be taken as unity. The fact that the two-dimensional gas equation holds for a *solid* unimolecular film, suggests that even in such films the molecules have freedom of movement; it may be noted in this connection that there is evidence that molecules in certain adsorbed layers are able to move within the bounds of the surface. This point should always be borne in mind when the state of an adsorption film is being envisaged. In a closely packed film the molecules are, of course, deprived of this freedom of movement

Built-up (Multimolecular) Layers.—An interesting consequence of the study of monolayers has been the formation of built-up films which may contain hundreds, or even thousands, of oriented molecular layers (K. B. Blodgett, 1935). A quantity of stearic acid is spread on a Langmuir trough (Fig. 257),

* The kinetic energy is assumed to be $\frac{1}{2}kT$ in each of two translational degrees of freedom (p. 329).

the underlying solution containing a small concentration, e.g., 3×10^{-4} molar, of a soluble salt of a bivalent cation, usually calcium or barium. The pH is adjusted to a little over 7 by means of sodium bicarbonate. The film is enclosed by a loop of silk thread, acting as a movable barrier, on the other side of which is placed a layer of a suitable oil, e.g., oleic acid, which maintains a constant surface pressure of about 29 dynes per cm. on the unimolecular film. The latter thus consists of an oriented layer of calcium or barium stearate molecules, depending on the nature of the ions present, under constant compression. If a clean strip of glass or metal is now inserted through the film, there is no deposition on the strip, but when it is withdrawn it is coated with a single layer of stearate molecules, the latter being oriented in such a manner that the polar portions are attached to the metal or glass, while the nonpolar hydrocarbon chains point outward. Upon inserting the strip once more, it now picks up another layer of molecules upon immersion, this layer being the mirror image of the first one, so that the polar groups are now outside. As the strip is removed it takes up a third layer which corresponds with the first. This procedure can be repeated several hundred, or even thousand, times until films producing visible interference effects are obtained. In this way the thickness of the built-up layer, and hence that of each monolayer can be calculated; the results are in excellent agreement with those derived from X-ray measurements. Further, electron diffraction studies show that the multilayers consist of hexagonal crystals, their axes being as expected, perpendicular to the plane of the film. It is of interest to note that if the glass or metal strip is initially coated with a film of ferric stearate, a molecular layer is also formed upon the first insertion, as well as upon withdrawal. The outside of the built-up film is now polar, whereas it is nonpolar when a clean strip is employed.⁹

ADSORPTION BY SOLIDS FROM SOLUTION

Adsorption and Interfacial Tensions.—The property possessed by charcoal of removing coloring matter from solution was known at least as far back as 1785 (T. Lowitz), and many observations have since been made showing that finely-divided powders are able to take up dyestuffs and other substances from solution; this property is another aspect of the phenomenon of adsorption. The function of the surface may be considered from two aspects: it may be regarded as acting in the same manner as in the adsorption of gases, i.e., by virtue of molecular attraction or of chemical forces, or it may be looked upon as a means for providing an interface of large area at which a solute capable of lowering the interfacial tension may accumulate. These two properties are undoubtedly related to each other, but the latter viewpoint is probably more useful; in any case other properties, less clearly understood, may also be involved. Very little information is available concerning the tensions at solid-liquid interfaces, but an approximate indication of the tendency for a given solute to be adsorbed at such an interface may be obtained by assuming that the tensions run parallel to those for air-liquid interfaces. There is no theoretical justification for this assumption, but failing any more definite information it may be employed as the only available guide, especially as experimental evidence appears to support the general validity of the parallelism between solid-liquid and air-liquid interfacial tensions. For example, dyestuffs are more readily adsorbed by charcoal from aque-

ous than from alcohol solutions; in fact the dyestuff adsorbed by charcoal from water may be readily extracted by means of alcohol. Water has a much higher air-liquid interfacial tension than has alcohol, and so the decrease of the surface tension of the aqueous solution with increasing concentration of the dyestuff is likely to be greater than for the alcoholic solution. According to the Gibbs equation, therefore, the adsorption at the air-liquid interface will be greater in aqueous than in alcoholic solution, and the same result is observed at the solid-liquid interface. Sometimes, however, the parallelism breaks down. Amyl alcohol has a greater influence than phenol in lowering the surface tension of water; nevertheless charcoal adsorbs phenol more readily from aqueous solution.

On account of the high surface tension of water, most substances are positively adsorbed from it. The relatively few cases of negative adsorption by solids, e.g., powdered silica or charcoal, from aqueous solutions occur with electrolytes; as seen above (p. 1205) these generally cause the air-liquid surface tension to increase. If blood charcoal is shaken with a solution of potassium chloride, for example, the concentration of the solution is found to be *greater* than the initial value; this negative adsorption of the solute is sometimes regarded as positive adsorption of the solvent. The simple application of the Gibbs equation is, however, insufficient to account for the observations concerning adsorption of electrolytes, for at high concentrations positive adsorption of the solute occurs from solutions of potassium chloride. One reason for this is the failure of the Gibbs equation to take into consideration electrostatic forces.

Adsorption and Concentration.—The extent of adsorption from solution increases with concentration, but it is probable that a limit is attained in adsorption by a solid surface from solution, just as is the case in the adsorption of gas by a solid and of a solute at the air-liquid interface. In the adsorption of acetic acid from aqueous solutions by both sugar charcoal and animal charcoal, the amount taken up reaches a maximum, varying with the nature of the adsorbent, which does not change on further increase of the concentration of the solution (G. C. Schmidt, 1911). Similar observations have been made on the adsorption of dyestuffs from aqueous solution. Results of this type suggest the formation of a unimolecular adsorbed layer, as in other types of adsorption, but such a conclusion must be accepted with reserve. There is evidence, as seen below, that in the adsorption of dyestuffs by charcoal and other finely divided substances a single molecular layer is not exceeded (F. Paneth, 1922), but on the other hand some workers are of the opinion that adsorbed films may be much thicker. For the present, therefore, the matter must be regarded as undecided.

Whatever the mechanism of adsorption from solution, it is certain that the extent depends mainly on the available surface of the adsorbent, although specific factors * are sometimes evident, especially with colloidal

* For example in dyeing wool and other materials simple adsorption is followed by chemical combination between the fiber and the dyestuff.

particles. The order in which a series of solutes are taken up from solution is therefore, in general, very much the same for different forms of charcoal, silica, and other finely divided adsorbents. The process of adsorption is almost invariably reversible and a definite equilibrium is reached in a short time dependent on the concentration of the solution and the quantity of adsorbent.

The experimental method employed is to shake a definite weight of the adsorbent with a known volume of the solution under investigation; partial evacuation may be necessary to remove air from the pores of the solid so that

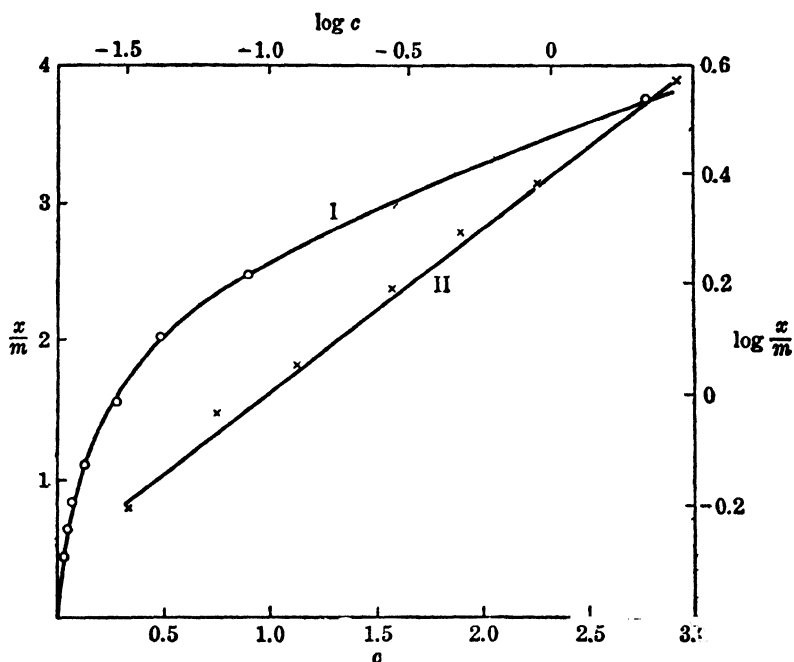


FIG. 259. Adsorption of acetic acid from aqueous solution

complete wetting may be ensured. After a short time the solid is allowed to settle and the clear liquid analyzed; this gives the equilibrium concentration of the solution. The amount of material adsorbed is calculated from the initial and final concentrations and the volume of solution employed.

The classical adsorption isotherm (p. 1196) in the form

$$\frac{x}{m} = kc^{1/n} \quad (33)$$

or

$$\log \frac{x}{m} = \log k + \frac{1}{n} \log c, \quad (34)$$

where c is the equilibrium concentration of the solution, applies over a considerable range of concentration in many instances. It appears to represent adsorption from solution more closely than it does adsorption of gases. The points in Fig. 259 are plotted from the data of H. Freundlich (1909) for the adsorption of acetic acid from aqueous solution by blood charcoal at 25° c.; curve I shows the variation of x/m with c , whereas II is the plot of $\log x/m$ against $\log c$. According to (34), II should be a straight line, and the agreement is seen to be satisfactory. The value of n is 2.35, and figures varying from about 2 to 4 have been obtained with charcoal and various substances adsorbed from aqueous solution. The extent of adsorption from solution decreases with increasing temperature, so that heat must be evolved in the process; there are, however, few quantitative data on the subject.

It is clear from curve I in Fig. 259 that for dilute solutions, as with gases at low pressure, the relative extent of adsorption is large. This is a matter of importance in analytical chemistry, since it renders difficult the removal by washing of the last traces of solutes adsorbed by precipitates. In order to minimize the effects of adsorption, precipitation should be carried out in hot solution, where the extent of adsorption is less, and the conditions should be adjusted so that the particles formed are relatively large and hence have a comparatively small exposed area. It should be noted, however, that large crystals sometimes contain impurities in the form of "inclusions."

Surface Area of Adsorbent.—If finely powdered lead sulfate is shaken with a solution containing a radioactive isotope, e.g., thorium-*B*, interchange of the isotopic atoms occurs and at equilibrium the ratio of the two isotopes in the solid will be equal to that in solution. In practice equilibrium is attained rapidly, and hence it is probable that exchange takes place only with the surface atoms, since there is insufficient time for penetration of the crystal lattice. It follows, therefore, that at equilibrium the ratio of lead to thorium-*B* in solution is equal to the ratio on the surface of the lead sulfate powder. The quantity of lead in solution can be determined analytically, and the amounts of thorium-*B* in solution and in the solid can be estimated by a suitable method which makes use of its radioactive properties. It is then possible to calculate the number of atoms of lead, and hence of molecules of lead sulfate, on the surface of a given weight of lead sulfate powder. From a knowledge of the approximate dimensions of the lead sulfate molecule the area of exposed surface can be estimated (F. Paneth, 1922-24). The same method has been extended to other insoluble lead salts and to the sulfates of calcium, barium and strontium, making use of the radioactivity of thorium-*B* or thorium-*X* which form isomorphous sulfates. By means of the artificial radioelements (p. 175), the same principle may be employed to ascertain the surface areas of insoluble salts of other elements.

Studies have been made of the maximum adsorption of dyestuffs, e.g.,

methylene blue, methyl green, naphthol yellow, etc., from solution by insoluble lead salts of known surface area (F. Paneth, 1924). The results indicated, as a general rule, the formation of an almost complete unimolecular layer of molecules lying flat on the surface. As a rough approximation, it appeared that 1 milligram of methylene blue covered 10^4 sq. cm. of surface, and this result was confirmed by using diamond powder as adsorbent, the exposed area being known quite accurately from measurements with a microscope. It has been suggested, therefore, that this generalization, or a similar one for other dyestuffs, may be used as a rough method for estimating surface areas of various adsorbing materials; its application would involve, of course, the assumption that a limit of adsorption is attained on completion of a unimolecular layer.¹⁰

Chromatographic Adsorption.—An important application of adsorption from solution is to be found in the methods of **chromatographic analysis**, the principle of which was first used by M. Tswett (1906), but not widely applied until R. Kuhn (1931) employed it in connection with his work on the carotenes. The procedure is based on the fact that a given solid may adsorb to varying extents the different constituents of a mixture present in a solution. If such a solution is poured down a long column packed with the powdered adsorbing solid, the material taken up by the upper layers will consist largely of the solute that is most readily adsorbed, together with lesser amounts of the others. The latter will be gradually adsorbed as they pass farther down the column. In this way a partial separation of the constituents of the mixture has been achieved. It may be greatly improved, that is, "developed," however, by allowing some of the pure solvent to flow slowly through the column. The less easily adsorbed solutes are thus removed from the upper parts of the column, but are reabsorbed at lower levels. The result is to spread the adsorption further down the column, while at the same time, provided the differences in the ease of adsorption of the substances are sufficiently marked, the adsorption of each solute is virtually restricted to a definite layer in the column. The most strongly adsorbed substance is at the top, while successive layers contain the other solutes in the order of their ease of adsorption.

Some of the earliest observations of this type were made with substances having different colors, and hence they appeared after separation as a series of colored bands; it is for this reason that the term "chromatographic adsorption" has been used for the procedure described. The solutes are generally organic compounds dissolved in acetone, petroleum ether, or other suitable solvent, and the adsorbing material consists of finely powdered alumina, magnesium oxide, charcoal, sucrose, etc. Upon completion of the separation, the column of solid may be pushed out of the containing tube and cut up into sections from which the separate solutes can be extracted. The individual layers can be readily seen when the substances are colored; in other cases use may be made of differences in their fluorescent behavior in ultraviolet light.¹¹

ELECTROKINETIC (ZETA) POTENTIAL

Electroosmosis.—When two portions of water are separated by moist clay or sand, and a difference of potential maintained, with a suitable source of E.M.F., by two metal electrodes placed one on each side of the clay or sand, a flow of water is observed from one side to the other (F.

Reuss, 1809; R. Porret, 1816). This phenomenon, known as **electroendosmosis** or **electroosmosis**, is almost invariably observed when an E.M.F. is applied to electrodes placed on either side of a porous diaphragm immersed in water or other liquid. If the liquid is initially at the same level on both sides of the separating membrane, the level will rise on one side and produce a difference of pressure until the rate of electroosmotic flow in one direction is exactly balanced by the flow in the opposite direction resulting from the excess pressure; a stationary state is thus attained which, as will be seen shortly, depends on the experimental conditions. On the other hand, if the levels are kept the same on both sides by means of an overflow tube, a continuous flow of liquid may be kept up as long as the E.M.F. is applied. A porous diaphragm is actually a mass of fine capillaries, and electroosmosis is found to occur also in capillary tubes, singly or in groups (G. Quincke, 1861).

The first quantitative measurements of electroosmosis were made by G. Wiedemann (1852, 1856); he found that the difference in pressure maintained between two sides of a porous diaphragm, of a given material, was proportional to the applied E.M.F., but was independent of the dimensions of the diaphragm. If the pressure is the same on both sides, the mass of liquid transported through a diaphragm in a given time is proportional to the current; for a given diaphragm material and given current it is independent of the length and cross section of the diaphragm. It was thought at one time that the flow of water was in the same direction as the current, but Quincke showed this was not always the case, and he was led to suggest that electroosmosis was due to the existence of electrically charged layers of opposite sign at the boundary between the liquid and the material of the diaphragm. The application of an E.M.F. must result in the displacement of the charged layers relative to each other, and since the liquid is free to move, whereas the diaphragm is not, a flow of liquid occurs; the direction of flow will depend on whether the liquid carries the positive or negative charge. This interpretation of electroosmosis also accounts for the fact, discovered by G. Quincke (1859), that if a liquid is forced through a porous material, e.g., a clay diaphragm or a capillary tube, a difference of potential is set up between the two sides; this is known as the **streaming potential**, which may be regarded as the reverse of electroosmosis. The forcible separation of the oppositely charged layers by pressing water through the membrane must, according to the theory, result in the observed difference of potential. Other phenomena due to the same fundamental cause, to which the general term **electrokinetic effects** may be applied, are the movement of particles under the influence of an applied E.M.F., to be considered later, and the production of a potential difference when particles fall through water, sometimes known as the **Dorn effect** (E. Dorn, 1880).

The Electrical Double Layer.—The concept of the existence of differently charged layers, or of the **electrical double layer**, at the solid-liquid boundary was extended by H. von Helmholtz (1879), who sug-

gested that an electrical double layer is generally formed at the surface of separation between two phases. By making the assumption that the double layer is virtually an electrical condenser with parallel plates no more than a molecular distance apart, he was able to treat the subject mathematically. According to the supposition of Helmholtz the potential gradient at the interface should be sharp, but modern views are in favor of a **diffuse double layer**, proposed by G. Gouy (1909) and supported by others. More recently O. Stern (1924) has shown that neither the sharp nor the diffuse double layer theory is alone adequate, and he has developed a view combining the essential characters of both. According to Stern the double layer is in two parts: one, which is approximately a single ion in thickness, remains almost fixed to the solid surface. In this layer, therefore, there is a sharp fall of potential. The second part extends some distance into the liquid phase and is diffuse; in this region thermal agitation permits the free movement of the particles, but the distribution of positive and negative ions is not uniform, since the electrostatic field at the surface will result in a preferential attraction of those of opposite sign. The result is a gradual fall of potential into the bulk of the liquid where the charge distribution is uniform. The condition at the solid-liquid boundary may be represented diagrammatically by Fig. 260, where the shaded portion represents the solid and the vertical broken line the extent of the fixed part of the double layer; the thickness is somewhat exaggerated in the diagram. If the potential of the solid is indicated by A and that of the bulk of the liquid by B , the fall of potential in between may occur in two ways (Fig. 260, I and II), depend-

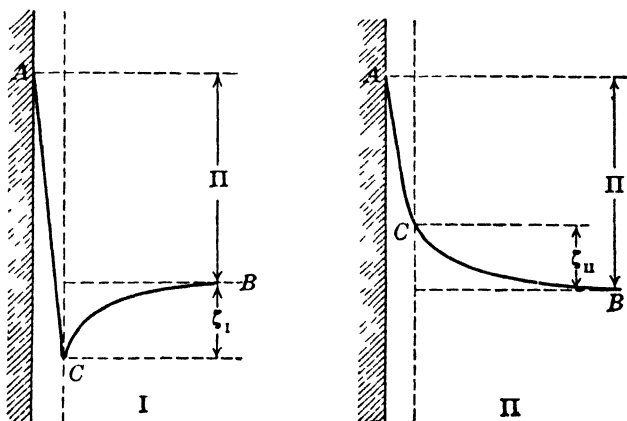


FIG. 260. Electrokinetic potential

ing on the characteristics of the ions or molecules present in solution which make up the outer portion of the fixed layer. In each case AC is the sharp fall of potential in the fixed part and CB the gradual change in the diffuse part of the double layer. The potential involved in electroosmosis

and allied phenomena is that between *C* and *B*, that is, between the fixed and freely mobile portions; this has been marked ζ in each diagram. The term **electrokinetic potential** was first suggested for this quantity, but in recent times there has been a tendency to use the noncommittal term **zeta-potential**, because of the symbol conventionally employed for its representation. Mention may be made here of the distinction between the electrokinetic potential and the thermodynamic, or reversible, potential existing between the solid and the bulk of the solution; the latter is represented by Π^* in Fig. 260, I and II. The two quantities are quite independent and must not be confused. The addition of capillary active substances, such as dyestuffs, or of ions of high valence has a marked effect on the zeta-potential at a glass-liquid surface, even to the extent of reversing the sign; the thermodynamic potential, as measured by the glass electrode (p. 995), is, however, hardly affected (H. Freundlich, *et al.*, 1920, 1925). The added substances do not apparently affect the values of the potentials at *A* and *B*, so that Π remains unaltered, but the potential gradient *AC* may be markedly altered because of changes in the outer portion of the fixed double layer.¹²

Quantitative Treatment.—In order that simple mathematical considerations may be applied to electrokinetic phenomena it is necessary to assume that the diffuse double layer is equivalent to an electrical condenser with parallel plates *d* cm. apart, each carrying a charge *e* per sq. cm.; the difference of potential between the plates may be taken as equal to ζ , the zeta-potential. If *D* is the dielectric constant of the medium between the hypothetical plates, then it is known from electrostatics that

$$\zeta = 4\pi ed/D. \quad (35)$$

This is the fundamental equation for the quantitative treatment of all types of electrokinetic phenomena.

When a liquid is forced by electroosmosis through the fine capillaries constituting a porous diaphragm, the rate of flow will be determined by two opposing factors, namely, the force of electroosmosis and the frictional force between the moving liquid layer and the wall. When the two forces are equal the flow will occur at a uniform rate. If *u* is the uniform velocity thus attained, then *u/d* may be taken as the velocity gradient in the double layer, since the velocity at one side, i.e., the wall, is zero, whereas the average value is *u* at the other side, i.e., in the moving liquid.† It follows, therefore, that the force due to frictional effects is equal to $\eta u/d$, where η is the coefficient of viscosity of the liquid. The electrical force causing electroosmosis is given by *Ee*, where *E* is the potential gradient across the membrane and *e* is the charge per sq. cm. at the boundary of movement; hence at the steady state

$$Ee = \eta u/d. \quad (36)$$

* The symbol Π is used here because a true single potential is implied, as distinct from the so-called "potentials" in Chapter XII, which are really E.M.F.'s.

† The actual velocity at this side of the double layer is less than *u* and increases toward the center of the tube; the case is simplified by supposing the system to involve a stationary wall and a liquid moving with uniform velocity.

Substituting the value of d given by this relationship in (35) it follows that

$$\zeta = \frac{4\pi\eta u}{DE}. \quad (37)$$

If q is the area of cross section of all the pores in the diaphragm,* the volume V of liquid transported electroosmotically per second is equal to qu , so that

$$u = V/q, \quad (38)$$

and hence

$$\zeta = \frac{4\pi\eta V}{qDE}. \quad (39)$$

In the case of electroosmosis through a single capillary tube, q is of course equal to πr^2 , where r is the radius of the tube, so that

$$\zeta = \frac{4\eta V}{r^2 DE}. \quad (40)$$

The strength I of the current is related to E by the expression $E = I/q\kappa$, where κ is the specific conductance of the liquid (cf. p. 889), so that (39) and (40) become

$$\zeta = 4\pi\eta\kappa V/DI. \quad (41)$$

The volume of liquid transported by electroosmosis should thus be proportional to the current strength, as found by Wiedemann (p. 1220).

As seen above, if liquid is allowed to accumulate on one side of a diaphragm, an excess hydrostatic pressure is set up which eventually counterbalances the electroosmotic flow. The simplest case is the flow through a single capillary tube, when Poiseuille's equation may be applied. The volume V of liquid of viscosity η flowing, per second, through a capillary tube of length l and radius r under a difference of pressure P is given by (p. 497)

$$V = \pi r^4 P / 8\eta l. \quad (42)$$

Substitution of this value for V in (40) gives

$$P = 2\zeta DEl / \pi r^2, \quad (43)$$

where P is the maximum pressure difference between the ends of the capillary tube as the result of electroosmosis. By assuming a porous diaphragm to consist of n capillary pores, each of radius r , it is a simple matter to show that exactly the same equation is applicable. This is in agreement with Wiedemann's observation that for a given diaphragm material, i.e., πr^2 is constant, the difference of pressure maintained is proportional to the applied voltage, i.e., to El , since ζ and D are unaffected by the dimensions of the diaphragm.

It is possible to utilize (43) to calculate the zeta-potential at the interface between glass and water by making use of the measurements on the

* It is assumed that the thickness of the double layer is negligible in comparison with the pore diameter.

TABLE 196. ZETA-POTENTIAL AT GLASS-WATER INTERFACE FROM ELECTROOSMOSIS IN CAPILLARY TUBES

r	El	h	ζ
0.0449 cm.	76 volts	0.0011 cm.	0.054 volt
0.0449	152	0.0023	0.053
0.0369	7,200	0.16	0.048
0.0369	11,500	0.25	0.046

difference of level of the water at the ends of a capillary tube to which an E.M.F. is applied. In Table 196 the first column gives the radius of the tube, r cm., the second the total voltage, i.e., El , and the third the difference in the levels of the water, h cm.; from these data ζ in volts,* in the last column, was calculated. The first two lines are from experiments by G. Quincke (1861) and the others by S. Terechin (1887). In view of the different conditions, i.e., radius of tube and voltage, and probably of the difference in the nature of the glass, the agreement of the results in the last column is good. The value of the electrokinetic potential at the glass-water interface is thus approximately 0.05 volt, the glass being negative relative to the water. It may be recorded here that measurements of various kinds, based on different types of electrokinetic phenomena, have shown that for most substances the zeta-potentials lie between +0.05 and -0.05 volt. The results cannot, of course, be exact since the equations do not take into account the diffuse nature of the solid-liquid boundary.

The velocity of a liquid flowing in a capillary tube varies with the distance x from the center of the tube, and is equal to $P(r^2 - x^2)/4\eta l$; the moving part of the double layer is at a distance $r - d$ from the center, and so its velocity u_d is given by

$$u_d = rdP/2\eta l, \quad (44)$$

the very small quantity d^2 being neglected in comparison with $2rd$. If one side of the double layer is forced past the other, as in the experiments on the streaming potential, the strength of the current I produced is

$$I = 2\pi r e u_d = \pi r^2 e d P / \eta l. \quad (45)$$

If κ is the specific conductance of the liquid used in the observations, then the actual conductance of the liquid in the capillary tube will be $\pi r^2 \kappa / l$, and if S is the streaming potential produced, it follows by Ohm's law that

$$\begin{aligned} S &= Il / \pi r^2 \kappa \\ &= edP / \eta \kappa, \end{aligned} \quad (46)$$

or substituting the value of ed from (35),

$$S = \zeta PD / 4\pi \eta \kappa, \quad (47)$$

so that the streaming potential is seen to be directly related to the zeta-potential. For a definite capillary diaphragm and liquid the streaming E.M.F. should be proportional to the driving pressure P ; this has been verified by R. A. Gortner (1931) with a diaphragm made of pulped filter paper.

From (41) and (47) it will be seen that V/I and S/P should be equal; it is

* See footnote, p. 1239.

of interest to record, therefore, that measurements by U. Saxén (1892) with a clay diaphragm, employing solutions of zinc, cadmium or copper sulfate, were found to be in agreement with this expectation. Some of the results are given in Table 197; in each case the quantities are in $\text{cm.}^{5/2} \text{ g.}^{-1/2}$ units. The close

TABLE 197. ELECTROOSMOTIC AND STREAMING POTENTIAL MEASUREMENTS

Solution	V/I	S/P
Zinc sulfate: $\frac{1}{2}$ per cent	0.360	0.352
Copper sulfate: 1 per cent	0.385	0.385
Cadmium sulfate: $\frac{1}{2}$ per cent	0.582	0.588

correspondence between these figures suggests that the equations deduced above, although they involve a simplifying assumption as to the structure of the double layer, may be used to give an indication of the magnitude of electrokinetic effects.

Influence of Ions on Electrokinetic Effects.—In the course of a comprehensive investigation of electroosmosis J. Perrin (1904) found that the magnitude, and even the sign, of the zeta-potential could be altered by the presence of ions in solution; hydrogen and hydroxyl ions are specially effective in many cases. The results in Table 198, for diaphragms made

TABLE 198. INFLUENCE OF ACID AND ALKALI ON ELECTROOSMOSIS

Carborundum		Naphthalene	
Solution	V ml./min.	Solution	V ml./min.
0.02 N HCl	+10	0.02 N HCl	+38
0.008 N HCl	0	0.001 N HCl	+28
0.0002 N HCl	-15	0.0002 N HCl	+3
0.0002 N KOH	-60	0.0002 N KOH	-29
0.002 N KOH	-105	0.001 N KOH	-60

of powdered carborundum * and of naphthalene, show the influence on the rate of endosmosis V , in ml. per min., of various solutions; the signs included indicate the charge on the diaphragm. In general, the zeta-potential is made more positive by acids, i.e., by hydrogen ions, and more negative by solutions of alkalis, i.e., hydroxyl ions; similar conclusions have been reached from measurements of the streaming potential.

If the hydrogen ion concentration of the solution is kept constant, the electrokinetic phenomena are found to be influenced by salts; the data recorded in Table 199 are from the work of J. Perrin (1904). The carbo-

TABLE 199. INFLUENCE OF SALTS ON ELECTROOSMOSIS

Carborundum in Alkali		Chromic Chloride in Acid	
Salt	V ml./min.	Salt	V ml./min.
—	-60	—	+90
0.1 M KBr	-14	0.1 M KBr	+33
0.002 M $\text{Ba}(\text{NO}_3)_2$	-26	0.001 M MgSO_4	+23
0.0002 M $\text{La}(\text{NO}_3)_3$	-18	0.0005 M $\text{K}_3\text{Fe}(\text{CN})_6$	+3

* According to F. Fairbrother (1924) *pure* carborundum gives somewhat different results.

rundum diaphragm in the presence of alkali was negatively charged, whereas that of chromic chloride had a positive charge in acid solution. These results suggest that the ion which is mainly effective in influencing endosmosis is the one with sign opposite to that of the diaphragm; this is clearly seen, for example, when comparing the results for barium and lanthanum nitrates on the negative carborundum diaphragm. Further, the higher the valence of the ion of opposite sign the greater is its effect in reducing the electroosmotic flow. It must not be imagined, however, that ions of the same sign as the diaphragm have no influence; this is shown by the fact that somewhat different results are obtained by changing the apparently ineffective ion, e.g., potassium chloride and ferrocyanide affect a negative diaphragm to different extents. The influence of the ion of the same sign as the diaphragm may be seen from the results of H. Freundlich and G. Ettisch (1925); in Fig. 261 are shown their

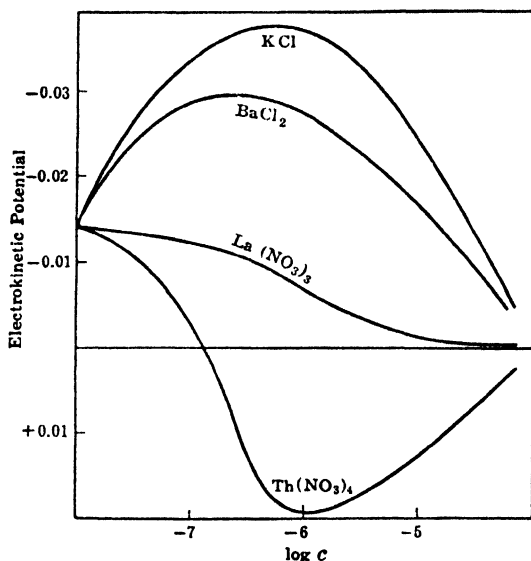


FIG. 261. Zeta-potentials in presence of salts

values for the zeta-potential at the glass-liquid interface determined from measurements of the streaming potential with a capillary tube. The initial increase in the negative electrokinetic potential on addition of potassium and barium chlorides suggests that the chloride ions are at first effective, but at higher concentrations the influence of the ions of opposite sign becomes evident. When positive ions of high valence, e.g., La^{+++} and Th^{++++} , are present the effect of these ions is so considerable that any influence of the anions must be very small in comparison. It will be observed that at high concentrations of thorium nitrate the zeta-potential tends to reverse its sign again because of the influence of the

negative nitrate ion. Similar observations have been made with cellulose diaphragms and aqueous solutions of various salts (R. A. Gortner, 1931).¹²

Electrolytes and the Zeta-Potential.—The results described can be explained in terms of the zeta-potential and the concept of the double layer at the solid-liquid interface. Substances like carborundum, sulfur and carbon, which do not ionize, are negatively charged in contact with water, and the addition of small amounts of a uni-univalent electrolyte tends to increase this charge. It is believed, therefore, that in these cases the negative zeta-potential is due in the first place to the firm attachment to the surface of hydroxyl ions from water, and possibly also of anions from the electrolyte. An equivalent number of positive ions, some closely held in the fixed part of the double layer and the rest in the diffuse portion, will remain in the liquid. The fall of potential from solid to the bulk of the liquid is shown diagrammatically in Fig. 262, I. If the electrolyte concentration is increased there will be a tendency for the cations to accumulate on the solution side of the fixed double layer, i.e., in the vicinity of the line XY in the diagram. It was considered at one time that the sole effect of this accumulation was to decrease the charge density e , but it now appears to be established that the thickness d of the double layer is decreased at the same time (cf. p. 1247, footnote). It follows from (35) that a diminution in either d or e , or both, will result in a decrease of zeta-potential,* as depicted in Fig. 262, II. The higher the valence of the cation the lower the concentration of the solution required to bring about sufficient change in the adsorption layer to produce a given effect (cf. p. 1246). If the positive ion concentration is made large, the sign of the electrokinetic potential may eventually be reversed (Fig. 262, III); this has been attributed either to the neutralization of the charge on the particle or, alternatively, to the diminution of the thickness of the double layer to such an extent that it eventually collapses and re-forms with the charges reversed.

Hydrogen and hydroxyl ions, large organic ions and ions of heavy metals, all of which are known to have a tendency to be held in the adsorption layer of a neutral substance, such as carbon, have a marked influence on the electrokinetic potential. It should be noted that in Fig. 262 the assumption has been made that the potential of the solid remains unchanged; this is not necessarily true, for the presence of strongly adsorbable positive ions will tend to make it more positive and this will favor still further the decrease and eventual reversal of the zeta-potential.

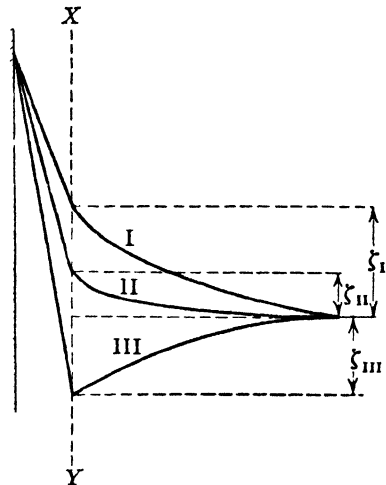


FIG. 262. Diagrammatic representation of the effect of ions on the zeta-potential

* If d decreases sufficiently ζ will decrease even though e increases; this is believed to occur in some cases.

If the material of the diaphragm is a substance capable of ionization, then the sign and magnitude of the zeta-potential may be determined by such ionization. For example, diaphragms of silica, tungstic and stannic acids, acid dyestuffs, soaps and glass, are negatively charged in contact with water, since in each case there is a tendency for a light cation, e.g., of hydrogen or alkali metal, to pass into solution leaving the complex anion to remain part of the negatively charged diaphragm. On the other hand, aluminum, and other basic, hydroxides and basic dyestuffs, which can take up protons from aqueous solution, form positively charged diaphragms. The influence of hydrogen and hydroxyl ions in these cases can be explained in terms of ordinary electrochemical equilibria. The addition of salts may result first in the preferential adsorption, on the solution side of the fixed double layer, of the ion having the same charge as the diaphragm, thus increasing the magnitude of the zeta-potential, e.g., glass and chloride ions (p. 1226), but it will be followed by the adsorption of the ion of opposite sign leading to the decrease, and even reversal of sign, of the double layer potential. The reversed potential continues to increase, but a still further increase in concentration of electrolyte may result in the re-adsorption of the anion and a decrease of potential, as with thorium nitrate.

It will be noted that in Fig. 261 as the concentration of electrolyte is increased the apparent zeta-potential tends in every case toward zero; the same result has been found in other cases. The exact significance of this fact is not yet clear; it may be due to the potential becoming very small, because of a decrease in thickness of the double layer, or it may be that the potential is appreciable but the observed electrokinetic effects are small owing to changes in the diffuse boundary layer and of the viscosity and dielectric constant of the liquid in that layer.

No satisfactory view has yet been expressed concerning the origin of the zeta-potential when the liquid is a nonconductor. It has been stated, however, that with such liquids electrokinetic phenomena are either entirely absent or else they are small and not well defined (J. Perrin, 1904). Nevertheless, zeta-potentials of appreciable magnitude have been measured. According to the studies of A. Coehn (1898, 1909), a substance acquires a negative charge when in contact with a material of higher dielectric constant; thus glass is positively charged in contact with turpentine, whereas sulfur, which has a lower dielectric constant, is negatively charged. Since the same generalization applies to the separation of charges resulting from friction, it was suggested that the electrokinetic potential and frictional electricity might have something in common. Considerable doubt has been cast, for various reasons, on the validity of Coehn's rule; it is certainly not of value when ionizing solvents are concerned, but it may have some use when the liquid is not an electrolytic conductor. The double layer then appears to involve either oriented polar molecules or an actual separation of electrons in some manner; as indicated above, however, the situation is obscure.

ELECTROCAPILLARY PHENOMENA

Electric Charge and Surface Tension.—It has been known for many years (W. Henry, 1801) that the properties of a mercury surface are affected by an electric charge, and the first comprehensive investigation of the subject was made by G. Lippmann (1873); his apparatus (Fig. 263),

consisted of a long tube (A) containing mercury, drawn out into a capillary (B), which is immersed in an aqueous solution. As a result of the interfacial tension between the latter and the mercury, the surface of separation appears in the capillary tube; the difference in level h between the upper and lower surfaces of the mercury is a measure of the interfacial tension. By connecting the mercury in the tube to one pole of a battery, and the mercury layer C to the other, a definite potential could be applied and the change of interfacial tension measured by finding the alteration in the height of the mercury column required to bring the surface in the capillary tube back to its original position. The mercury has initially a positive potential with respect to the solution, but as it is made gradually more negative by means of the applied E.M.F., it is found that the surface tension at first increases, reaches a maximum and then decreases; the curve showing the variation of surface tension with the applied potential (Fig. 264) is called an **electrocapillary curve**.

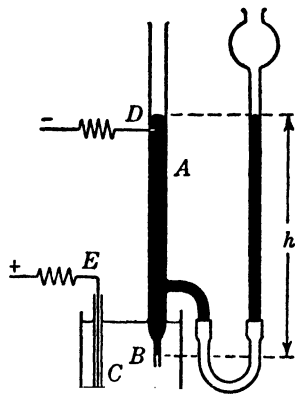


FIG. 263. Apparatus for the study of electrocapillarity (Lippmann)

The general form of the curve is independent of the nature of the electrolyte: it may be accounted for in the following manner. A double layer exists at the interface between mercury and solution, with the former carrying the positive charge; this may result either from the adsorption of ions present in the electrolyte or from the tendency for the mercury to send electrons into the solution. The charges on each side of the double layer will tend to repel each other, so that the work required to increase the area of the interface, and hence the interfacial tension, is relatively small. If the mercury is made more negative by the applied E.M.F., the double layer potential is decreased, and the lower charge density at the surface must be accompanied by an increase of surface tension. When the potential of the double layer is completely annulled the interfacial tension will be a maximum, for further increase of the applied voltage results in the formation of a double layer of opposite sign to the original. The theory of the electrocapillary curve has been treated from the thermodynamic standpoint by Lippmann, J. Willard Gibbs and others, and it has been shown that the variation of interfacial tension with applied E.M.F., i.e., $\partial\gamma/\partial E$, at constant temperature, pressure and composition, should be equal to the charge density at the interface. Since $\partial\gamma/\partial E$ is zero at the maximum of the curve, it follows that the double layer must then be uncharged.

Influence of Added Ions.—Although the general shape of the electrocapillary curve is always the same, the presence of various substances in

the electrolyte results in some modification. With a series of potassium salts, for example, the left-hand, i.e., ascending, branch, may vary somewhat in position, but the right-hand, i.e., descending, branch is unchanged. Certain anions, e.g., iodide and sulfide, produce a very marked effect, as shown qualitatively by the broken line in Fig. 264; the maximum surface tension is lowered and a more negative potential is required for its attainment. While the mercury is positively charged it will attract

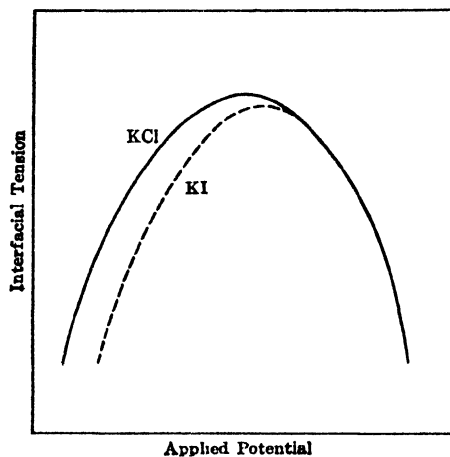


FIG. 264. Electrocapillary curves

anions toward the interface, and the more readily these are adsorbed, e.g., iodide and sulfide, the more will they lower the interfacial tension. At the same time, the accumulation of charge on the negative side of the double layer will mean that in order for its potential to be annulled, and the maximum of the curve attained, a larger negative voltage will have to be applied to the mercury. Once the maximum of the electrocapillary curve is passed, the mercury acquires a negative charge and now the anions are repelled instead of being adsorbed; the descending portion of the curve will thus be independent of the nature of the anions present in the solution.

As is to be expected from these arguments, strongly adsorbed cations, e.g., tetra-alkyl ammonium ions, do not affect the left-hand branch, but alter the right-hand portion of the electrocapillary curve. These substances lower the maximum interfacial tension, as they must do if they are adsorbed, and shift the requisite potential to more positive regions. Non-ionic compounds that are strongly adsorbed, e.g., the higher alcohols, have little influence on either branch of the curve, but they depress the maximum, in agreement with anticipation.¹⁴

THE COLLOIDAL STATE

Introduction.—In the course of his investigations on diffusion, T. Graham (1861) found that substances could be differentiated according to the rates at which their solutions passed through a parchment membrane; compounds which diffused readily, e.g., sugar, salts, etc., generally existing in crystalline form, he called **crystalloids**, but those diffusing at a very slow rate, e.g., gelatin, albumin, glue and silicic acid, were given the name **colloids** (Greek: *glue-like*). Although the study of colloids as a special branch of chemistry may be said to have begun with Graham, solutions of a colloidal nature were known at a much earlier date. For example, colloidal gold had been known for many years before it was prepared independently by M. Faraday (1857); colloidal silicic acid was known to T. Bergman (1779) and to J. J. Berzelius (1833), the latter having also obtained colloidal solutions of arsenious sulfide (1844). The most important early investigations on the subject were made by F. Selmi (1845–47) who described “pseudo-solutions,” or what are now known as colloidal solutions, of silver chloride, sulfur and prussian blue.

It was realized by Graham, and more clearly by others in later years, that the distinction between crystalloidal and colloidal *substances* was not rigid; many crystalline substances, e.g., sulfur, graphite and sparingly soluble salts, can be obtained in colloidal solution. Further, X-ray studies have shown that particles in colloidal systems are often truly crystalline in character. The sodium salts of long-chain fatty acids, i.e., the soaps, have properties associated with colloids in aqueous solution, but are more like crystalloids when dissolved in alcohol; gum arabic behaves in a somewhat similar manner. It is evident, therefore, that it is impossible to divide substances into crystalloids and colloids, and consequently the latter term is now applied to a particular *state* in which many substances can exist. Instead of employing the term “colloidal substance,” it is the practice to refer to the “colloidal state” or to a “colloidal system.”

General Properties of Colloidal Systems.—In a true solution, such as one of urea or sucrose, the particles of solute distributed in the solvent are believed to be of molecular size. On the other hand, a suspension or emulsion (p. 1275) contains particles large enough to be visible to the naked eye, or at least in the microscope, distributed in a liquid medium. Between these two extremes are to be found the colloidal systems; that is to say, the essential characteristic of the colloidal state is the existence of particles which are larger than molecules but not large enough to be seen in the microscope. It is obviously impossible to draw a distinct line between colloidal solutions and true solutions on one side, and between colloidal solutions and suspensions on the other; nevertheless colloidal systems have certain properties which, as a general rule, place them in a category distinct from the other systems mentioned.

The upper limit of size of particles in the colloidal state is usually taken

as the lower limit of microscopic visibility, viz., about $0.2\ \mu$, whereas the lower limit is about $5\ m\mu$.* The latter figure approximates to the diameter of certain complex molecules of high molecular weight, e.g., proteins, starch and rubber; it is not surprising, therefore, that solutions of such materials show colloidal properties, and it is easy to understand why Graham applied the term "colloid" to substances of this type. It is evident, however, that the *colloidal state* is not necessarily restricted to these complex compounds, but it is one capable of adoption by any sparingly soluble substance.

Dialysis and Filtration.—The best filter papers will retain only particles larger than about $1\ \mu$ in diameter, and so it is seen that colloidal solutions pass through filter paper like true solutions. By the use of special devices it is possible, however, to separate colloidal particles from those of molecular dimensions. The procedure known as **dialysis**, first employed by Graham, utilizes the fact that substances in true solution, provided the molecules are not too large, can pass through a parchment membrane, while colloidal particles are retained. It must not be imagined that the dialyzing membrane acts merely as a sieve depending on the size of its pores; this is no doubt responsible to some extent for the selective behavior, but differences in diffusion and adsorption of the small and large particles also play an important part.

Various forms of dialyzer may be used to remove substances in true solution from a colloidal system. In the simplest a sheet of parchment paper is fixed over the end of a glass cylinder, or over a wide metal ring. The experimental solution is then poured into the resulting container and the whole is suspended in an outer vessel containing water which is renewed either periodically or continuously. On standing, the dissolved substance passes through the membrane and is washed away. Instead of the familiar drum-shaped dialyzer, long tubes of parchment, known as "sausage skins," and seamless thimbles of parchment paper, similar in shape to Soxhlet thimbles, have been employed. Since parchment paper offers considerable resistance to the passage even of molecules, dialysis with this material is slow; in recent work, therefore, it has been largely replaced by collodion, particularly for work on a small scale. Dialyzers of cellophane are also employed for a similar purpose. In **electrodialysis** the removal of electrolytes is facilitated by means of an electric field; the colloidal solution is placed between two dialyzing membranes, with pure water in a compartment on each side. By means of electrodes in these compartments an **E.M.F.** is applied which causes the ions to migrate out of the colloidal solution.

The separation of molecular from colloidal particles may also be carried out by **ultrafiltration** (H. Bechhold, 1907); by impregnating filter papers with collodion, or with gelatin subsequently hardened by formaldehyde, the pores can be made small enough to retain particles of colloidal dimensions. The solution is poured into the filter, and the passage of the liquid is accelerated by pressure or suction. The size of the pores depends

* The symbol $m\mu$ is used for a millimicron, i.e., $10^{-6}\ \mu$, $10^{-7}\ \text{cm.}$ or $10\ \text{\AA.}$

on the particular paper employed and the concentration of the solution used to impregnate it; it is thus possible to obtain a series of graded ultrafilters by means of which a colloidal solution may not only be purified from dissolved material, but it may be separated into fractions containing particles of different sizes. An approximate estimate of the size can be obtained from a knowledge of the dimensions of the pores, determined from the pressure required to force air or water through them. It must be strongly emphasized that the results are very approximate, since, as with dialysis, pore size is by no means the only factor determining whether a given particle will pass through an ultrafilter or not.

Optical Properties.—Although colloidal particles are too small to be retained by filter paper or to be seen in the microscope, their presence can be made evident by optical means. If a strong beam of light is passed through a medium which is optically clear, that is, it contains no particles appreciably larger than molecules, the path of the light cannot be easily detected; if particles of larger size are present, however, they scatter the light and the beam is rendered visible. Most true solutions are optically clear, but colloidal solutions, as M. Faraday (1857) first showed with colloidal gold, scatter the light, thus indicating the presence of particles which are larger than molecules but too small to be separated by filtration. The phenomenon of the scattering of light by particles was studied by J. Tyndall (1869) and hence it is generally called the **Tyndall effect**. In the device known as the **ultramicroscope**, invented by H. Siedentopf and R. Zsigmondy (1903), the scattered light in the Tyndall beam can be observed in a microscope; in this way the presence of individual particles in a colloidal solution is made evident as flashes of scattered light. It should be pointed out that the actual particles may be too small to be visible to the microscope; it is possible, however, to see the light scattered by them.

Further Properties of Colloids.—Another important property of many colloidal systems is their sensitivity to small quantities of electrolytes. It is known that substances in true solution may sometimes be "salted out" by the addition of a relatively large amount of an electrolyte, but with many colloidal systems the presence of very small quantities of ionizable substances causes the particles to coagulate so that a visible precipitate is readily formed.

Colloidal systems are often called **sols** to distinguish them from true solutions; sometimes a prefix is added, e.g., as in **aquasols** or **alcosols**, to indicate the nature of the liquid involved. The terms **precipitation**, **coagulation** or **flocculation** are employed for the formation of a visible solid* by various means, e.g., addition of electrolyte, evaporation or cooling. The solid was originally described as a **gel**, but this term is now generally restricted to those cases in which the whole system sets to

* Graham used the term "pectinization," later abbreviated to "pectization" (Greek: *curdles*), but this is now rarely encountered.

a semi-solid state without any supernatant liquid initially, although some liquid may be extruded later (p. 1263). The process of forming a gel is called **gelation**. Some writers employ the term gel to include gelatinous precipitates, e.g., aluminum and ferric hydroxides, formed from colloidal systems, whereas others refer to these as **coagels**. When the word gel is used in this wider sense, the special case in which no liquid is formed is referred to as the formation of "jelly," as is the case with table jellies.

It will become evident, as the treatment of colloidal solutions is developed, that the essential character of these systems is dependent on the large ratio of area to volume for the particles involved. In a true solution the system consists of one phase only and there is no true surface of separation between the molecular particles of solute and solvent; colloidal solutions, however, are two-phase systems and for each particle there is a definite surface of separation between it and the liquid medium. At this interface characteristic surface properties, such as adsorption and double layer potentials, are evident. When the total surface area is relatively small, as with large particles in suspensions, the influence of these properties is not marked, but in colloidal systems the area is so large that they play an important part. For example, the total surface area of 1 cc. of material when it exists as a cube of 1 cm. side is 6 sq. cm.; when divided into cubes of 10^{-6} cm. side, which approximates to many colloidal systems, the total area of the same volume of material is 6×10^6 sq. cm. It is obvious, therefore, that the function of the surface will be of great importance.

Disperse Systems.—The two phases involved in a colloidal system may be distinguished by the terms **disperse phase**, for the phase forming the particles, and **dispersion medium**, for the medium in which the particles are distributed (Wolfgang Ostwald, 1907); the whole may be referred to as a **disperse system**. It was assumed above that the dispersion medium was a liquid, but it may be solid or gaseous; similarly, the disperse phase may be solid, liquid or gaseous, leading to a number of possible disperse systems. Smokes and dusts, for example, consist of solid particles dispersed in a gaseous medium, whereas in fog, mist and cloud the disperse phase is liquid; these are sometimes called **aerosols**. Many minerals contain fine inclusions of gas, liquid or solid dispersed in a solid medium, and ruby glass consists of colloidal particles of gold dispersed in glass. A foam is a two phase system involving gas dispersed in a liquid medium. Although many of these disperse systems are of considerable technical interest, they will not be discussed; most attention will be paid to those in which the dispersion medium is a liquid. If the disperse phase is a solid and the particles relatively large, a suspension is obtained, but if the disperse phase is liquid the result is an **emulsion**; with smaller particles the familiar colloidal systems are formed.

Lyophobic and Lyophilic Systems.—Colloidal solutions with a liquid as dispersion medium can be divided roughly into two categories; it was

thought at one time that they corresponded to the presence of particles of solid or liquid, respectively, and were consequently given the names "suspensoids" and "emulsoids." This classification is now regarded as misleading, and the one in common use is to divide colloidal systems into **lyophobic** (Greek: *liquid hating*) and **lyophilic** (Greek: *liquid loving*) types; when water is the dispersion medium the terms "hydrophobic" and "hydrophilic" are generally employed. The essential differences between the two types of sols are given below. Another distinction between

LYOPHOBIC SOLS

1. Surface tension similar to that of dispersion medium.
2. Viscosity similar to that of medium.
3. Small quantities of electrolytes cause precipitation.
4. The particles are easily detected in the ultramicroscope.
5. The particles migrate in one direction in an electric field.

LYOPHILIC SOLS

1. Surface tension often lower than that of dispersion medium.
2. Viscosity much higher than that of medium.
3. Small quantities of electrolytes have little effect, but large amounts may cause salting out.
4. The particles cannot be readily detected in the ultramicroscope.
5. The particles may migrate in either direction or not at all in an electric field.

lyophobic and lyophilic systems is that on evaporation or cooling, the former give solids which cannot generally be readily reconverted into sols by reversing the physical change; lyophilic sols are, however, reversible in this respect and the solids obtained are true gels, i.e., of the jelly type. Typical examples of hydrophobic sols are those of metals, sulfur, sulfides and silver halides; and sols of gums, starches, proteins and soaps provide instances of hydrophilic systems. As in most cases of colloidal behavior it is not possible to draw a sharp line of demarcation between lyophobic and lyophilic sols, since some, particularly those of hydroxides, often show intermediate properties. The difference between the two types of colloids is not due to the particles being solid or liquid, as was at one time thought, for a mercury sol, in which the particles are undoubtedly liquid, is lyophobic in character like other metal sols, e.g., gold, with definitely crystalline particles. It may be significant, however, that substances which most readily form lyophilic sols are those of high molecular weight; each particle, therefore, consists of a small number of molecules and in some instances possibly of only one large molecule. It is consequently readily understood why lyophilic sols behave to some extent like true solutions, e.g., with regard to influence of electrolytes and evaporation. In view of the differences in behavior of the two main types of sol it is convenient to examine their properties separately.

LYOPHOBIC SOLS

Optical Properties.—The Tyndall effect resembles fluorescence in appearance, but actually the phenomena are quite different. The light scattered in the Tyndall beam is polarized and the distribution of intensity

among the wave lengths is determined to a great extent by the size of the particles. Lord Rayleigh (1871, 1899) showed that for electrical insulators the relationship between the wave length λ of the light and the intensity I of scattering was

$$I = k(v^2/d^2\lambda^4), \quad (48)$$

where k is a constant, v is the volume of the particle, and d is the distance from the particle to the observer. Since the intensity is inversely proportional to λ^4 , it is clear that radiations of short wave length will be most intense in the scattered light; this accounts for the fact that many colorless or pale colored sols have a bluish Tyndall beam. The constant in Lord Rayleigh's equation contains a term involving the difference between the refractive indices of the particles and the dispersion medium; when these are equal there should be no scattering, and this may account for the small Tyndall effect observed with lyophilic sols and with some lyophobic sols.

The Ultramicroscope.—Although particles smaller than about 0.2μ cannot be seen directly in the best ordinary microscopes, they are still able to scatter light and it is this scattered light which, as previously recorded, is visible in the ultramicroscope; the presence of particles as small as 5 to 10 $m\mu$ diameter can be made evident in this manner. It is essential, however, as in the Tyndall effect, that the refractive index of the disperse phase and dispersion medium should be appreciably different. The principle of the slit ultramicroscope, of Siedentopf and Zsigmondy, is illustrated diagrammatically in Fig. 265. If a narrow beam of parallel,

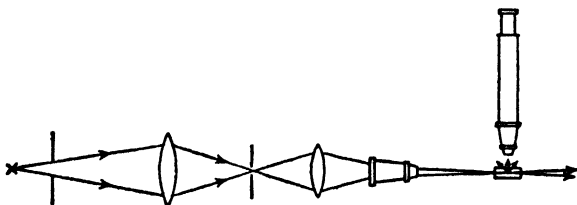


FIG. 265. Slit ultramicroscope

or slightly convergent, light is passed at right angles to the direction of the microscope through a cell on which the instrument is focused, the field will appear completely dark provided no particles are present in the cell. On the other hand, if it contains a colloidal solution the particles will scatter the light, some of which will pass vertically into the microscope; each particle thus appears as a small disc of light on a dark background. With a powerful source of light, e.g., an arc lamp, this arrangement is superior to any other for the observation of colloidal particles, but fairly satisfactory results can be obtained in a simple manner by the use of a special sub-stage microscope condenser giving dark-field illumination.

These devices are so designed that light is normally totally reflected at the cover glass and no light enters the body of the instrument; if particles are present in the liquid placed on the microscope slide, however, some of the light is scattered and so becomes visible (Fig. 266).

Brownian Movement.—As is to be expected, because of their small size, colloidal particles are seen in the ultramicroscope to display vigorous Brownian movement (p. 255), and the equations already deduced (pp. 257, 262), on the assumption that the motion is caused by molecular impacts, have been found to hold just as well for particles of colloidal dimensions as they do for microscopic granules. As already seen in Chapter IV, measurements with colloidal particles of known radius can be used for determining the Avogadro number.

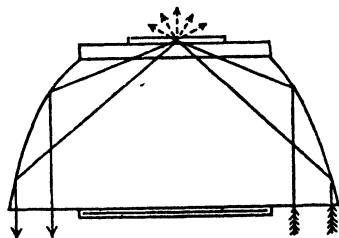


FIG. 266. Dark-field condenser

Determination of Particle Size.—The size of colloidal particles may obviously be obtained from observations on the Brownian movement and of the distribution under the influence of gravity, provided the Avogadro number is known, by reversing the procedure referred to above. This principle has been developed in the “ultracentrifuge” and used to determine the sizes and particularly the “molecular weights” of lyophilic colloids (see p. 1256). An approximate indication of the dimensions of the particles in a colloidal system is given by the use of ultrafilters (p. 1232), but the most reliable procedure is to employ the slit ultramicroscope. If the weight of material in a given volume of sol is known, and the density of the particles is assumed to be the same as in the macroscopic state, the total volume of the disperse phase can be calculated; to determine the volume of each particle it is only necessary, therefore, to count the number of particles in a given quantity of the sol. The counting is carried out by placing the sol, diluted to a known extent if necessary, in the cell of a slit ultramicroscope and noting the number of particles visible; the observation is repeated many times and the average taken. The length and breadth of the field of vision are measured with the aid of an eye-piece micrometer, and the depth is determined by rotating through 90° the slit which transmits the beam of light; from these dimensions the volume of the sol containing the number of particles counted can be evaluated. The requisite information for calculating the volume of a single particle is now available, and if it is assumed to be spherical the radius can be determined. Another, more approximate, method is to measure in the ultramicroscope the average distance l between two particles in a given sol. Then l^3 is the volume which, on the average, contains a single particle; there are consequently, on the average, $1/l^3$ particles in unit volume of sol. If m is the mass of the disperse phase contained in unit volume,

and ρ is the macroscopic density, then the total volume of the particles may be taken to be m/ρ , and hence the volume of each particle is ml^3/ρ .

Shape and Color of Particles.—The characteristic colors shown by many sols are related in some degree to the particle size; in the course of coagulation, for example, the color of a gold sol changes from red to violet and then to blue. The wave length of the light adsorbed by a particle is approximately proportional to its radius; hence the larger the particle the shorter the wave length of the light transmitted, i.e., the color approaches the violet end of the spectrum. The shape of the particle, however, has an important influence on color, and the difference between red and blue gold sols is believed to be determined mainly by the spherical form of the particles in the red and the disc-like particles in the blue sols. Some information concerning the shapes of colloidal particles has been obtained by the study of the Tyndall scattering of individual particles and of flowing sols. If the particles are spherical, the scattering of light will be the same in all directions, but if they are rod-like they will tend to orient themselves in a flowing liquid and so give different effects along and at right-angles to the direction of flow.

Electrical Properties: Electrophoresis.—When a hydrophobic sol is placed in an electric field the particles move definitely in one direction or another, indicating that they are electrically charged with respect to the dispersion medium (S. E. Linder and H. Picton, 1892). The phenomenon of the migration of colloidal particles under the influence of an applied E.M.F. (cf. p. 1220) was called **cata-**

phoresis, although the term **electrophoresis**, now coming into general use, is preferable. Sols of metallic oxides and hydroxides, and basic dyestuffs are usually positively charged, while sols of metals, sulfur, metallic sulfides, acidic hydroxides and acid dyestuffs carry a negative charge (cf. p. 1228). The velocity of migration and the sign of the charge carried by the particles can be determined by macroscopic or ultramicroscopic methods.

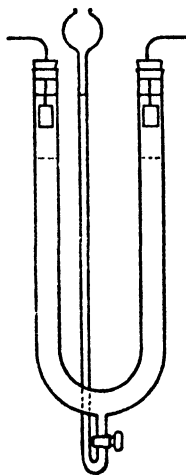


FIG. 267. Measurement of electrophoresis (Burton)

In the macroscopic method the apparatus shown in Fig. 267 (F. E. Burton, 1906) is convenient. The lower part of the U-tube contains the sol covered by the pure dispersion medium, e.g., water, in which dip platinum electrodes connected to a source of E.M.F. If the boundary between the sol and the pure liquid is sharp and visible, its rate of motion, which is equal to the average speed of the particles, can be determined. Special methods, e.g. causing the surface to fluoresce in ultraviolet light, are sometimes employed to render the boundary visible. The potential gradient is known from the applied E.M.F. and the dimensions of the apparatus, and hence the velocity of the particles under a fall of potential of 1 volt per cm. may be calculated. In the other method, a single particle is observed in

the ultramicroscope; a special microscope slide, to which electrodes are fixed, is used. On account of the electroosmotic movement of the liquid at the edges of the slide, measurements made on particles at the edge or in the middle will be in error; a particle at about one-fourth the distance between the two edges should thus be observed.

In accordance with the previous discussion of electrokinetic phenomena (p. 1220), electrophoresis may be readily explained by the presence of an electrical double layer at the interface between the disperse phase and the dispersion medium, and the consequent presence of a zeta-potential. When an E.M.F. is applied the positive and negative portions of the double layer must be displaced relative to each other, and since the particles in a sol are free to move they will migrate in the electric field. By treating the double layer surrounding a particle as a simple condenser, a relationship identical with equation (37), viz.,

$$u = \zeta DE / 4\pi\eta, \quad (49)$$

is obtained, where u is the observed velocity of the particle and u/E is the **mobility**, i.e., the velocity for a potential gradient of 1 volt per cm. More exact treatment by P. Debye and E. Hückel (1924) indicates that the numeral 4 in the equation is strictly applicable to cylindrical particles only, and that it should be replaced by 6 for spherical particles. It is not experimentally established, however, that the shape of the colloidal particles has any influence, and so (49) will be adopted here.

Some of the data collected by H. Freundlich, on the speeds of migration and the corresponding zeta-potentials* for various aquasols, are recorded in Table 200; the sign preceding the potential indicates the

TABLE 200. MOBILITIES AND ZETA-POTENTIALS OF PARTICLES IN WATER

Disperse Phase	Mobility	Zeta-Potential
Arsenious sulfide	2.2×10^{-4} cm./sec.	-0.032 volt
Prussian blue	4.0	-0.058
Gold	4.0	-0.058
Platinum	3.0	-0.044
Silver	2.4	-0.034
Ferric oxide	3.0	+0.044

charge carried by the particle. In almost all sols studied the particles have mobilities of about $2-4 \times 10^{-4}$ cm. per sec., so that the zeta-potentials, in agreement with previous findings (p. 1224), are about 0.03 to 0.06 volt. It is significant that particles of quartz 1μ in diameter, oil globules of 2μ diameter and even air-bubbles 0.1 mm. in size migrate in an electric field with approximately the same speed as do colloidal particles, so that the interface potentials are all of the order of 0.05 volt. The speed of a chloride ion, under a potential gradient of 1 volt per cm., is 6.8×10^{-4}

* In utilizing (49) to calculate the zeta-potential in volts from the rate of migration under a potential gradient of 1 volt per cm., it is necessary to divide both ζ and E by 300 (cf. p. 6).

cm. per sec., and hence colloidal particles migrate with a velocity comparable with that of electrolytic ions. The electrical conductance of a sol is quite low,* however, because the number of current carrying particles is very small compared with that in a solution of an electrolyte at an appreciable concentration. The larger charge carried by the colloidal particles is not sufficient to compensate for their smaller number.

Just as ions of sign opposite to that of a diaphragm alter the zeta-potential curve and affect electroosmosis, so ions influence the velocity of migration of colloidal particles in an electric field. E. F. Burton (1909) has determined the concentrations of three potassium salts necessary to reduce the mobility of the positively charged particles in a copper oxide sol from 2.5×10^{-4} to 2.0×10^{-4} cm. per sec.; the results are given in Table 201. It is evident that the anion is effective in decreasing the

TABLE 201. INFLUENCE OF VALENCE OF IONS ON MIGRATION OF COPPER OXIDE PARTICLES

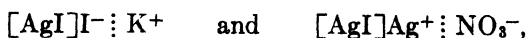
Electrolyte	Concentration to Reduce Velocity
Potassium chloride	0.115 mole/liter
Potassium sulfate	0.049
Potassium ferricyanide	0.0022

velocity and the zeta-potential, as is to be expected for a positive sol, and the influence of the ion increases markedly with increasing valence. In some cases the addition of a relatively large amount of a trivalent ion will change the sign of the charge carried by the colloidal particle; this has been observed with aluminum chloride when added to gold and silver sols, and with ferric chloride added to platinum sols. It will be seen shortly that at intermediate concentrations of electrolyte, when the zeta-potential is small, the sol is unstable and precipitation occurs.¹⁵

Origin of the Charge.—The electric charge carried by a colloidal particle is of fundamental importance, since without it colloidal systems would be very unstable. The presence of electrical double layers of the same sign surrounding all the particles results in their mutual repulsion, so that they do not approach sufficiently closely to coalesce. It should be noted that an increase in size of the particles would lead to a decrease of total area, and hence to a decrease of free energy of the system; union of the particles would thus be expected to occur were it not for the influence of the electric charges. It has been frequently observed that small quantities of electrolytes are associated with colloidal systems, and that if they are removed by persistent dialysis, or in other ways, the sol becomes unstable and the particles grow in size and are precipitated. It is probable, therefore, that in many cases traces of ions present in the sol are responsible for the charge and stability of the colloidal particle. It is of interest that reasonably stable sols of gold and silver can be prepared by the Bredig method of passing an electric arc or spark between

* Colloidal electrolytes (p. 1266) are, of course, exceptional; these resemble lyophilic sols.

electrodes of the metals under water (p. 1274) provided the latter contains about $5-10 \times 10^{-4}$ g. equiv. of alkali per liter. Further, if a dilute solution of a silver salt is added to excess of dilute potassium iodide a negatively charged sol of silver iodide is obtained, while if the dilute iodide solution is added to excess of the silver solution a positively charged sol results; if the two solutions are mixed in equivalent amounts the sol is unstable and complete precipitation occurs (A. Lottermoser, 1905). The stability and charge of the negative sol formed in the presence of excess of iodide ions may be explained by the adsorption of these ions on to the particle side of the double layer at the surface, whereas the positive sols are similarly stabilized by the adsorption of silver ions, since the latter are in excess. As a general rule, ionizable substances appear to have a preference for taking up ions which are common to them, thus indicating a natural tendency for the crystal to extend its own space lattice. This accounts for the existence of the two oppositely charged silver iodide sols, which may be represented as



where $[\text{AgI}]$ implies a particle of silver iodide and the dotted line indicates approximately the limit of the fixed part of the double layer* (p. 1227). Although only one kind of ion is shown to the left and one to the right of the dotted line, this must be understood as implying only that this particular ion is present in excess.

Ferric oxide sols have also been obtained both positively and negatively charged (F. Powis, 1915). If ferric chloride solution is added to a large volume of hot water and then dialyzed, a positively charged sol is obtained. Taking into consideration the tendency for a particle to extend its crystal lattice and the fact that the concentration of ferric ions is relatively large, the structure of the colloidal unit may be written as $[\text{Fe}_2\text{O}_3 \cdot x\text{H}_2\text{O}]\text{Fe}^{+++} \vdots 3\text{Cl}^-$, implying that the positive charge, and stability, is to be attributed to the ferric ions which are firmly attached to the particle. The inclusion of $x\text{H}_2\text{O}$ means that the ferric oxide is hydrated, but the extent and nature of the hydration, viz., chemical or physical, is not known. If dilute ferric chloride solution is run slowly into excess of dilute sodium hydroxide with constant shaking, a negatively charged sol is obtained, probably to be represented as $[\text{Fe}_2\text{O}_3 \cdot x\text{H}_2\text{O}]\text{OH}^- \vdots \text{Na}^+$, the hydroxyl ion acting as the stabilizer. It is not impossible that chloride ions may function in the same manner, just as hydrogen ions might aid the stability of the positive sol, but the structures suggested are based on the belief that the particle will prefer to be stabilized by an ion which is present in its own lattice. The so-called ferric oxide sols always contain

* The colloidal particle together with its double layer is sometimes called a "micelle"; this term was first used by C. von Nägeli (1879) to designate the equivalent in colloidal solution of the molecule in true solution. The term is now employed particularly in connection with colloidal electrolytes (p. 1267).

some chlorine, probably present in the particle as a basic chloride, but this does not affect the main argument concerning the origin of the charge.

Another possibility has arisen from a consideration of observations with stannic oxide sols; if stannic chloride is poured into water a precipitate of the oxide is obtained which, after washing with water, can be peptized either by hydrochloric acid to give a positively charged sol, or by potassium hydroxide to give a negative sol. To explain these results R. Zsigmondy (1915) suggested that the stannic oxide reacts with the hydrogen and hydroxyl ions giving stannic and stannate ions, respectively, and that these are adsorbed on the particle side of the double layer; thus,



This view is in harmony with the general idea, partly expressed in the tendency for a common ion to act as stabilizer, that surface molecules adjust themselves so as to make the transition from one phase to another as gradual as possible.

Sulfide sols are generally stabilized by hydrogen sulfide. It has been found, in fact, that suspensions of freshly precipitated sulfides, e.g., of arsenic and cadmium, may be peptized† by the passage of hydrogen sulfide; if the latter is completely removed precipitation occurs (S. W. Young and W. R. Goddard, 1917). Since the sols are negatively charged it is evident that the sulfide ion plays the predominant part in determining stability, and the simplest suggestion for the structure of the colloidal particle would be $[\text{As}_2\text{S}_3]\text{S}^{--} \vdots 2\text{H}^+$. Alternatively it is possible that the hydrogen sulfide first reacts with arsenious sulfide to produce $\text{H}_2\text{As}_2\text{S}_4$, or other thioarsenious acid, the ions of which are responsible for the double layer; thus, $[\text{As}_2\text{S}_3]\text{As}_2\text{S}_4^{--} \vdots 2\text{H}^+$.

The sols of the noble metals, silver, gold and platinum, appear always to have negative charges.* Positive sols of other metals, e.g., lead, bismuth and iron, have been described, but these may well be oxide or hydroxide sols, which are generally more stable when positively charged. The metallic sols mentioned are prepared by the arc or spark method, referred to above, in the presence of minute amounts of hydroxyl ions preferably, or sometimes of chloride, bromide or iodide ions. It is generally agreed that the charge and stability of the sols obtained in this manner are to be ascribed, in the first place, to the presence of hydroxyl ions. According to W. Pauli (1926) these react with the metal to form complex ions on the surface. If the conditions are such that no anions are present, then a complex anion may be formed as a result of interaction between the metal and water. The work with platinum has shown that the acid $\text{H}_2\text{Pt}(\text{OH})_6$ is present in solution after the electric discharge is passed, and it is probably an important factor in determining the stability of the sol; the particle may then be depicted as $[x\text{Pt} \cdot y\text{PtO}_2]\text{Pt}(\text{OH})_6^{--} \vdots 2\text{H}^+$.

* The addition of positive ions of high valence, e.g., Fe^{+++} or Al^{+++} , can yield a positive sol (p. 1245), but this is probably a mixed sol.

† The term "peptize" (Greek: *digest*), proposed by Graham, refers to dispersion into colloid form.

The inclusion of an oxide, e.g., PtO_2 , in the formula is justified by the view that some direct oxidation of the surface of the metal must always occur. That it may be an essential part of a metallic sol is suggested by the fact that silver sols are unstable in the presence of ammonia which is able to dissolve silver oxide.

In addition to the various ways described above, whereby a colloidal particle may acquire its charge, there is the possibility, previously considered in connection with the zeta-potential (p. 1228), of direct ionization of the material of the particle. This is probably of importance with certain acidic and basic dyestuffs; the former ionize to yield hydrogen ions leaving a negatively charged particle, whereas the latter take up protons and so form a sol with positive particles. Although self-ionization may sometimes determine the sign of the zeta-potential, it is doubtful whether it is often the cause of the stability of colloidal systems. If it were effective, then shaking a freshly precipitated metallic hydroxide with pure water should produce a sol; actually this does not occur, the addition of a little free alkali being required as peptizing agent. The hydroxyl ions are either adsorbed on the hydroxide and remain as such to give a negative charge to the particles, or else they may react, as would probably be the case with aluminum hydroxide, to produce another negative ion, e.g., $[\text{Al}_2\text{O}_3 \cdot x\text{H}_2\text{O}]\text{AlO}_2^-; \text{Na}^+$. This is equivalent to the explanation of the origin of the two types of stannic oxide sol given above.¹⁶

Precipitation by Electrolytes.—Traces of electrolytes are essential to the stability of sols, at least in water, but somewhat larger amounts cause the particles to aggregate and form a visible precipitate. The phenomenon of precipitation, flocculation or coagulation, as it is variously called, was observed by T. Graham (1861) and studied in some detail by H. Schulze (1882–83), by S. E. Linder and H. Picton (1895, 1905), and by W. B. Hardy (1900). In comparing the precipitating effects of different ions for a given sol it is essential that the experiments should be carried out under exactly comparable conditions; for example, different results may be obtained if the electrolyte is added very slowly or almost instantaneously. The concentration of the sol also has an important influence. Nevertheless, by standardizing the technique and allowing a prescribed time to elapse, it is possible to determine whether a solution of a given concentration produces precipitation or not. As a result of work of this kind it is evident that the ion which is effective in bringing about coagulation is the one of opposite sign to that of the colloidal particle, and that the coagulating power increases considerably with increasing valence of the ion; these conclusions concerning the effect of sign and valence are included under the title of the **Schulze-Hardy rule**, which is the analogue of Perrin's rule for electroosmosis (p. 1226).

The results in Table 202 show the minimum concentrations, in millimoles per liter, of various electrolytes required to cause precipitation of a positively charged sol, viz., ferric oxide (H. B. Weiser, 1920), and of a negative sol, viz., arsenious sulfide (H. Freundlich, 1910). In each case

TABLE 202. PRECIPITATING EFFECTS OF VARIOUS IONS

Ferric Oxide (Positive) Sol			Arsenious Sulfide (Negative) Sol		
Electro-lyte	Anion Valence	Min. Conc.	Electro-lyte	Cation Valence	Min. Conc.
KCl	1	103	NaCl	1	51
KBr	1	138	KNO ₃	1	50
KNO ₃	1	131	$\frac{1}{2}$ K ₂ SO ₄	1	63
KBrO ₃	1	31	HCl	1	31
K ₂ CrO ₄	2	0.325	MgSO ₄	2	0.81
K ₂ SO ₄	2	0.219	BaCl ₂	2	0.69
K ₂ C ₂ O ₄	2	0.238	ZnCl ₂	2	0.68
K ₃ Fe(CN) ₆	3	0.096	AlCl ₃	3	0.093

the result of increasing the valence of the ion of opposite sign to that of the charged particles is apparent, whereas a change in the ion of the same sign has only little influence. The precipitating effect of a tervalent ion is here 600 to 1000 times as great as that of a univalent ion.* An examination of Table 202 shows that there are some partial departures from the strict valence rule, e.g., potassium bromate on ferric oxide sol, and hydrochloric acid on arsenious sulfide sol; much more striking deviations are, however, known. The concentration of crystal violet, which yields a univalent cation, necessary to precipitate the arsenious sulfide to which Table 202 refers is only 0.16 millimole per liter, and morphine and strychnine hydrochlorides bring about precipitation at 0.4 and 0.5 millimole per liter, respectively; the univalent ions are here more effective than other bivalent ions. It should be noted that the ions which behave in this exceptional manner are those which are usually strongly adsorbed on surfaces. Less striking, but nevertheless important, discrepancies in the Schulze-Hardy rule may be seen in Table 203 giving the flocculating con-

TABLE 203. FLOCCULATING CONCENTRATION OF POTASSIUM SALTS FOR ARSENIOS SULFIDE SOL

Potassium Salt	Minimum Concentration	Potassium Salt	Minimum Concentration
Citrate	>240	Sulfate	63
Acetate	110	Chloride	50
Formate	86	Nitrate	50

centrations, in millimoles per liter, of various potassium salts for the arsenious sulfide sol referred to above. In spite of the fact that the potassium ion is the active one, the precipitating effect of the various salts is quite different; these results show that the anions must have some influence. Deviations of other types have also been observed; for example, the same concentration of calcium, barium and aluminum chlorides is required to precipitate a negative stannic oxide sol prepared by peptiza-

* These figures have no fundamental significance, for the relative amounts of ions of different valence required to cause precipitation vary with the concentration of the sol.

tion with alkali. Behavior of this kind, where ions of different valences have the same coagulating effect, is anomalous and is partly connected with chemical reaction. Many exceptions to the Schulze-Hardy rule are to be found in the literature and consequently it should be taken as a rough guide only, although probably in the majority of cases the rule is, at least approximately, applicable.

An unusual type of anomalous precipitation, analogous to the effect of ions of high valence on the charge of metal sols (p. 1240), has been observed. The addition of 0.083 millimole of ferric chloride to a negative platinum sol causes coagulation brought about by the ferric ions, but if 0.333 millimole is added rapidly to a fresh portion, a fairly stable positive sol is obtained, and only when the concentration is increased to 16.33 millimoles does precipitation occur again, now presumably by the action of the chloride ion. Similar results have been obtained with aluminum chloride and a silver sol, with thorium nitrate and a gold sol, with strychnine nitrate and arsenious sulfide sol, and with sodium phosphate and a positive ferric oxide sol. The particles in the positive sols obtained in this way probably contain a certain quantity of the hydroxide of the added ion.

Mechanism of Precipitation.—It has been seen that in general the presence of ions of opposite sign to that of the colloidal particles results in a decrease in the numerical value of the zeta-potential, and electrophoretic experiments have shown that sols have a critical value for this potential, about 0.02 to 0.03 volt, below which they cease to be stable. It appears that when the double-layer potential is decreased below the critical value, the repulsion between approaching particles is reduced to such an extent that those colliding with a certain velocity can join together; in this way coagulation will occur. The phenomenon of anomalous precipitation is readily understood, for addition of electrolyte in excess of a certain amount will not merely alter the zeta-potential, but will produce a system with a potential greater than the critical value and of opposite sign. The sol is thus stabilized, although further addition of electrolyte will eventually reduce the zeta-potential again and permit flocculation to occur.

There are still two important matters requiring consideration: in the first place, it is necessary to account for the change in the electrokinetic potential resulting from the addition of ions, and secondly, the influence of valence on the precipitating effect of an ion must be explained. A point of view, covering both these aspects of the problem, which has gained much support because of its simplicity, has been proposed by H. Freundlich (1910). It is assumed that the decrease of zeta-potential, leading to precipitation, is brought about by the adsorption of the ion of sign opposite to that of the colloidal particle.* If this is the case then electrically equivalent amounts of ions of different valences should be adsorbed; that is to say, the quantities of uni-, bi-, and ter-valent ions

* This may be due to decrease of charge density or thickness of the double layer, or both (see p. 1227).

adsorbed should be in the ratio of 3 to 1.5 to 1. The amount adsorbed may be related to the concentration of the solution by the classical isotherm, viz., $x/m = kc^{1/n}$, and making the assumption that all ions are adsorbed to the same extent, that is k and n are the same, it follows that

$$kc_1^{1/n} = 3.0, \quad kc_2^{1/n} = 1.5, \quad \text{and} \quad kc_3^{1/n} = 1.0,$$

where c_1 , c_2 and c_3 are the concentrations of uni-, bi-, and ter-valent ions in the solution necessary to cause the *adsorption* of the relative amounts 3, 1.5 and 1, respectively. It follows, therefore, that

$$c_1:c_2:c_3 = 3^n:1.5^n:1^n,$$

and if n has the reasonable value of 6, the ratios are 729 to 11.4 to 1; these are considerably different from the amounts actually adsorbed, and are in fair agreement with experiment (cf. Table 202). The same conclusion may be reached qualitatively from an examination of Fig. 268,

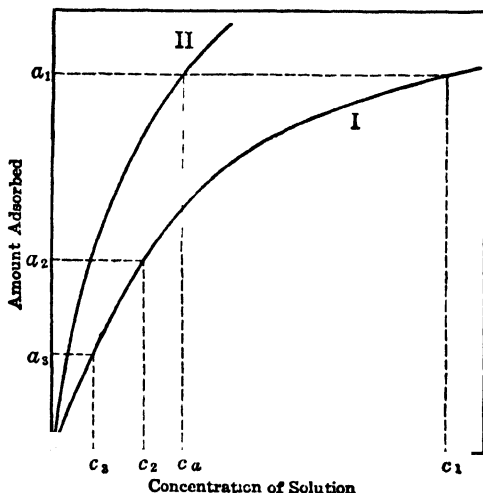


FIG. 268. Adsorption of ions of different valence

in which curve I shows a typical adsorption isotherm (cf. Fig. 259): a_1 , a_2 and a_3 are the electrically equivalent amounts of uni-, bi-, and ter-valent ions which must be adsorbed: the requisite concentrations of the solutions are then c_1 , c_2 and c_3 . It is evident that the ratio of c_1 to c_3 is much greater than that of a_1 to a_3 , so that the tervalent ion is a very effective precipitating agent. The arguments given are based on the assumption that all ions are equally adsorbed, which is of course not true; actually, large organic ions or those of high valence tend to be adsorbed more readily than simple inorganic ions. Suppose, for example, that curve II represents the adsorption isotherm for strongly adsorbed ions; the concentration of a univalent ion in the solution necessary to bring

about precipitation will then be c_a , which is much less than for the simple univalent ion c_1 . In this way the marked coagulating power of morphine and strychnine hydrochlorides, mentioned above, as well as the deviations of a smaller order of magnitude, can be explained. The influence of ions with the same sign as the colloidal particles (Table 203) can be interpreted by assuming that certain ions, e.g., citrate, are preferentially adsorbed and by increasing the zeta-potential confer additional stability on the system; it is necessary, therefore, for more of the ion of the opposite sign to be adsorbed than would otherwise be required before the zeta-potential is reduced to the critical value.

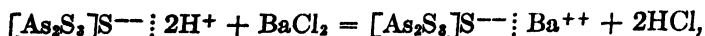
According to the theory of the double layer described on p. 1221, the ions which are effective in precipitation are adsorbed on the solution side of the fixed part of the double layer. Consider, for example, an arsenious sulfide sol peptized by a small amount of hydrogen sulfide, the structure of the particle being represented as $[\text{As}_2\text{S}_3]\text{S}^{--} : 2\text{H}^+$, and suppose a solution of barium chloride is added. Since the barium ion is the active precipitating ion, it must presumably replace, to a considerable extent, the hydrogen ions already in the double layer, viz., $[\text{As}_2\text{S}_3]\text{S}^{--} : \text{Ba}^{++}$. The hydrogen ions will then pass into the bulk of the solution which should thus become acid. Further, as a result of the relatively high concentration of barium ions remaining in solution, the charge density and thickness of the double layer will decrease and so the electrokinetic potential will diminish (p. 1227).¹⁷

The possibility has been suggested that a colloidal particle may be regarded as an ion of high valence and treated by the interionic attraction theory (p. 956). If this is so the quantity which has been called the electrokinetic potential may be regarded as the effective potential at the ionic particle due to its own charge and to the opposing effect of the surrounding "counter-ions" (or "gegen-ions") which constitute the ionic atmosphere. The potential [cf. equation (121), p. 958] depends essentially on $e^{-\kappa r}$, where κ , the reciprocal of the radius of the ionic atmosphere, is proportional to the square root of the ionic strength of the medium; the greater the latter, therefore, the smaller will be $e^{-\kappa r}$ and consequently also the equivalent of the zeta-potential.* An increase in concentration of ions in solution, especially if they are of high valence, will therefore result in an increase in the opposing potential produced by the ionic atmosphere, and hence will have the effect of facilitating precipitation. It is possible that the valence rule might be interpreted in this manner without recourse to the adsorption isotherm, but this is not certain.¹⁸

Whichever point of view is adopted, that is, either adsorption into the solution side of the fixed double layer, or increase in the opposing potential of the "ionic atmosphere," the structure of the colloidal particle in the presence of added electrolytes will be represented virtually in the same manner. The case cited above, i.e., addition of barium chloride to an

* The thickness of the ionic atmosphere decreases with increasing electrolyte concentration, and this may correspond to the postulated decrease in thickness of the double layer (see p. 1227).

arsenious sulfide sol, may be written



and it has in fact been found that when sufficient electrolyte is added to cause flocculation, the precipitate contains some barium salt; free hydrochloric acid is produced in the solution during the addition of the barium chloride. When other ions are used as coagulants these are found to be attached to the precipitate. It was at one time considered that equivalent amounts of different ions were thus carried down irrespective of their coagulating power; such is approximately the case for the flocculation of arsenious sulfide sol with a few ions, but it is far from being generally true. This is not surprising since equivalent quantities of different ions will not necessarily have the same effect on the thickness of the double layer, and the critical zeta-potential for precipitation may vary with the nature of the electrolyte.

Adsorption Indicators.—The application of the phenomena of adsorption to the detection of the end points of precipitation titrations was first made by K. Fajans (1923), and the results have been extended by him and by I. M. Kolthoff (1927 *et seq.*) who suggested the name **adsorption indicators**. If silver nitrate solution is added gradually to one of sodium chloride in the presence of a suitable dyestuff with a colored anion, e.g., fluorescein or eosin, the precipitate of silver chloride is quite normal until the solution contains a slight excess of silver ions, when the silver chloride takes up the colored anion from solution. The change in appearance is very marked and so the end-point of the titration is readily detected. Adsorption indicators have been employed in the titration of halides, cyanides and thiocyanates with silver nitrate, and of ferrocyanides, borates and phosphates with lead acetate. To explain the results, e.g., in the silver-chloride titration, it has been suggested that in the presence of excess of chloride ions negatively charged particles, represented by $[\text{AgCl}]\text{Cl}^-$, are formed; these cannot attach the negative dye anions to themselves to any appreciable extent. As soon as the silver ions are in excess, however, the particles of precipitate become positively charged, viz., $[\text{AgCl}]\text{Ag}^+$, and these immediately adsorb the colored anions, presumably on the solution side of the fixed part of the double layer. It is possible, as Kolthoff has pointed out, that there is an "ion-exchange" equilibrium between the chloride and dye ions in solution and on the precipitate. As long as the former are in excess in solution they will also predominate on the solid, but when their concentration is reduced to a very small amount, i.e., at the equivalence point, the colored anions will occupy most of the available parts of the surface.¹⁹

Mutual Action of Sols.—When two sols with charges of the same sign are mixed, no special phenomena are observed, provided no chemical reaction occurs, but the mixing of sols carrying opposite charges can result in mutual precipitation. From the comprehensive experiments made

by W. Biltz (1904) it appears that a sol behaves in this respect very much like an electrolyte. For example, when a positive ferric oxide sol was added to a negatively charged arsenious sulfide sol no appreciable flocculation occurred until a certain amount of the former had been added, when both sols precipitated. If a much larger quantity of the positive sol was added at one time, no flocculation took place but the arsenious sulfide was now stabilized by the ferric oxide as a positive sol. Mutual precipitation of oppositely charged sols is a general phenomenon; the coagulating power of a series of positive oxide sols is found to fall very approximately in about the same order for a number of negative sols. The cause of the flocculation is probably electrical interaction between the ions constituting the double layers of the respective sols.

Sensitization and Protection.—If a very small amount of a hydrophilic is added to a hydrophobic sol, it is sometimes observed that the latter has become more sensitive to precipitation by electrolytes; this **sensitization** appears to be partly due to adsorption by the hydrophobic particles of the oppositely charged hydrophilic substance, although other factors are undoubtedly operative. If the hydrophilic sol has the same charge as the hydrophobic particles, or if a relatively large quantity of the former of opposite sign is added, however, the hydrophobic sol becomes very much less sensitive to the precipitating action of electrolytes; the latter sol is in fact **protected** by the former. The protective action of various lyophilic substances was expressed by R. Zsigmondy (1901) in terms of the **gold number**: this is defined as the weight, in milligrams, of dry material from which the lyophilic sol is prepared, which when added to 10 ml. of a standard gold sol (0.0053 to 0.0058 per cent) is just sufficient to prevent the color change, from red to blue, on the addition of 1 ml. of a 10 per cent solution of sodium chloride. The smaller the gold number the greater is the protective action of the particular lyophilic substance. Some of the results of Zsigmondy are quoted in Table 204; the reciprocals of the gold

TABLE 204. PROTECTIVE ACTION OF LYOPHILIC SOLS

Sol	Gold Number	Reciprocal
Gelatin	0.005–0.01	200–100
Sodium caseinate	0.01	100
Hemoglobin	0.03–0.07	33–14
Gum arabic	0.15–0.25	7–4
Sodium oleate	0.4–1.0	2.5–1
Gum tragacanth	2	0.5
Potato starch	25	0.04

numbers are included as they give a comparative indication of the protective action. The lyophilic substances are in some manner able to confer their property of stability to electrolytes on the lyophobic sols. The simplest interpretation, which is widely employed, is that the protective colloid forms an adsorbed layer completely surrounding the particle of the lyophobic sol, so that it behaves virtually as one of lyophilic sub-

stance (H. Bechhold, 1904). Further inquiry shows that this explanation meets with difficulties. The amount of protective agent is often insufficient to cover the particles of lyophobic sol, and even if it were, the mechanism of the action is not easy to understand when it is remembered that the protecting substance itself consists of particles comparable in size with those being protected. It is reasonably certain that the particles of the two sols are united in some manner, but the exact state of affairs is not clear.

The remarkable protective action of salts of protalbic and lysalbic acids, obtained by the alkaline hydrolysis of albumin, has been utilized for the preparation of sols, particularly of metals, possessing exceptional stability. These sols may be evaporated to dryness, and the addition of water containing a little alkali will regenerate the sol, even after the lapse of several years (C. Paal, 1902 *et seq.*). An unprotected sol would be irreversibly coagulated on evaporation, and would in any case be relatively unstable. It is of interest to note that the stannic chloride formed in the reduction of a gold salt by stannous chloride is partially hydrolyzed, and the resulting stannic acid acts as a protective agent for the colloidal gold produced in this manner; the solid precipitated from this sol, known as Purple of Cassius (1685), is readily peptized by dilute aqueous ammonia.²⁰

LYOPHILIC SOLS

The chief lyophilic substances, e.g., starch, proteins, etc., are naturally occurring and their colloidal properties are probably of biological significance; it is for this reason that many of the investigations have been made by biochemists and physiologists. The study of lyophilic sols is very complicated, and the results are often difficult to understand; nevertheless, some of the general conclusions are definite enough to be stated.

Viscosity.—One of the most striking differences between lyophobic and lyophilic sols is the high viscosity of the latter. A difficulty is encountered at the outset in the treatment because different results are often obtained by the use of different experimental methods. Some workers are of the opinion that lyophilic sols do not obey Poiseuille's law, and that the apparent coefficient of viscosity varies with the rate of flow through a capillary tube. On the other hand, H. R. Kruyt (1925) has pointed out that if attention is paid to the construction of the viscometers employed and allowance is made for possible changes in the nature of the sol, such as gelation, Poiseuille's law holds provided the particles are spherical. In spite of the difficulties approximate viscosity data are available from which important conclusions can be drawn.

The viscosity of a system, such as a sol, consisting of a disperse phase and a dispersion medium has been considered by A. Einstein (1906), who derived an equation which may be expressed in the form

$$\eta_{sp.} = \frac{\eta}{\eta_0} - 1 = 2.5\phi, \quad (50)$$

where η is the viscosity coefficient of the system, η_0 that of the solvent, and ϕ is the ratio of the volume of the disperse phase to that of the whole system. Since ϕ is approximately proportional to the concentration of the sol, the specific viscosity should also be proportional to the concentration. It is not difficult to show that (50) is of the same form as the general equation (61), p. 501, with α equal to zero; it is probably applicable to spherical molecules only. Since few lyophilic substances, e.g., agar and proteins, are spherical, they do not satisfy (50); although the specific viscosity is approximately proportional to the concentration, the ratio is not in agreement with the expected value.

In spite of the quantitative failure of (50), the results indicate that for lyophilic sols ϕ is relatively large; the high viscosity of such sols may thus be explained by the large volume of the dispersed particles. This volume is considerably larger than that of the corresponding dry material, and hence the conclusion is to be drawn that the particles of the lyophilic sol are extensively solvated; the high degree of solvation thus accounts for the high viscosity. It is perhaps significant in this connection that solids like gelatin and agar, which form lyophilic sols readily in water, are able to take up large quantities of this liquid to form jelly-like masses (cf. p. 1259).

The Electroviscous Effect.—The addition of a small quantity of an electrolyte to a lyophilic sol produces a very marked decrease of viscosity; for example, 2 millimoles per liter of an electrolyte with a univalent cation, e.g., a potassium salt, diminishes the viscosity of an agar-agar sol by 20 per cent. Experiments with different salts have shown that it is the ion with sign opposite to that of the particles of the sol which is effective, and the influence increases greatly with increasing valence of the ion. It appears, therefore, that the decrease in viscosity is associated in some manner with a decrease of the double-layer potential. The small amount of electrolyte added cannot affect appreciably the solvation of the particles, and so it is probable that the zeta-potential is one of the factors, in addition to the effective volume of the disperse phase, determining the viscosity. The portion of the viscosity due to the charge on the particles is called the **electroviscous effect**. This property of lyophilic colloids is likely to prove of importance in the study of nonaqueous sols.

Surface Tension and Foaming.—The surface tensions of many lyophilic sols are lower than that of pure water, although the effect is apparently not universal, since starch and pure gelatin do not diminish the tension at the air-liquid interface. Where the lowering does occur the colloidal material must be adsorbed at the interface, according to the Gibbs equation; it is for this reason that different results are often obtained for surface tension measured by static, e.g., capillary tube, and dynamic, e.g., vibrating jet, methods, for in the latter the surface is being continually renewed and the equilibrium concentration is not attained.

Lyophilic sols produce foams very readily, and this is generally attributed to the lowering of surface tension. This may be true in some

instances, but it is now known that any substance capable of producing an appreciable change, increase or decrease, of surface tension will facilitate the formation of a foam. In other words, foam production is helped by the presence of a surface at which the concentration is different, either smaller or larger, than in the bulk of the liquid.

Sign of Electric Charge.—Particles of lyophilic sols migrate in an electric field, but the direction of movement is often, especially with proteins and related substances, very sensitive to changes in the hydrogen ion concentration. Although the electrophoretic method may be employed (see below), it is not often convenient because the particles are difficult to detect in the ultramicroscope; further, since the sols are colorless the boundary between them and water is not easily visible to the eye. Fortunately, the influence of ions on the viscosity may be used to determine the sign of the charge carried by the particle; where both this and the electrophoretic method have been employed the same conclusions have been reached. The charge carried by a protein or other particle of amphoteric nature is governed by the particular function, acidic or basic, effective in the given environment. If the solution is relatively acid, the lyophilic substance will behave as a base and by taking up the protons will become positively charged, but in alkaline solutions there will be a tendency for protons to be given off and a negative sol will result. The negative charge of agar-agar may also be attributed to a form of self-ionization, for it appears that it consists to a great extent of a complex sulfuric ester of calcium which on ionization would give a negatively charged residue. The question of the charges carried by soaps, which have a similar origin, will be considered later.

Electrophoresis: The Tiselius Method.—The electrophoresis of lyophilic sols has been studied by both microscopic and macroscopic methods (p. 1238), special devices, e.g., ultraviolet light, being used in the latter case to render visible the surface of separation between the sol and water. An important development in this field is the modification of the macroscopic procedure by A. Tiselius (1937 *et seq.*) so as to permit the separation and possible identification of proteins in biological fluids. The essential part of the apparatus is a U-shaped glass cell of rectangular cross section, consisting of two parts, one of which can be made to slide over the other. The lower part is filled with the lyophilic sol under investigation, usually in a suitable buffer medium, while the upper part contains the buffer solution alone. By sliding the latter into position, so as to complete the U-form, there is obtained a sharp boundary between the sol and the pure buffer medium. The two limbs of the U-cell are connected to two large electrode vessels, thus permitting an E.M.F. to be applied; as a result, electrophoresis occurs and photographic devices are used to render the movement visible. In the "shadow" method the boundary appears as a dark line on a light background. If two or more different substances, having different electrophoretic mobilities, are present in the sol, two or more boundaries will form in the course of time, each moving steadily ahead of the succeeding ones. From the rate of motion, the mobilities of the various particles can be determined and often identified. If one boundary leads the succeeding one by

a sufficient distance, it is possible to effect a separation by sliding the two parts of the cell relative to each other. In the "scanning" technique, which is a modification of the foregoing, each boundary appears as a peak on the photograph; the area under any peak is proportional to the amount of protein causing that peak present in the sol. This form of the electrophoretic procedure thus permits of a quantitative, as well as a qualitative, analysis of the lyophilic system. It should be noted that two proteins having the same mobility would always give one boundary, and hence are not separable by the electrophoretic method.

Stability of Lyophilic Sols.—Small amounts of electrolytes undoubtedly diminish the zeta-potentials of lyophilic sols, as is implied by their effect on the viscosity; nevertheless no coagulation occurs. It is even possible to obtain lyophilic sols without charge, i.e., in the isoelectric state, when the particles do not show electrophoresis, yet the sol is quite stable. It is apparent, therefore, that some factor in addition to the charge must play a part in determining the stability of such sols. At one time it was thought that the particles in a lyophilic sol were actually large single molecules and so were in a sense real solutions; this may be true in certain instances (see p. 1258), but it does not account for all the types of lyophilic sols. The work of H. R. Kruyt (1922–28) has provided strong evidence that hydration is a fundamental factor in determining the stability of such sols; two aspects of the subject may be described. If considerable amounts, viz., 50 per cent or more, of alcohol or acetone are added to an agar-agar sol, striking changes are observed; the viscosity decreases, the Tyndall effect is increased, and relatively small amounts of electrolytes cause precipitation, there being a marked valence influence. The flocculating effect of ions is also observed if the alcohol or acetone is added after the electrolyte. It appears that the lyophilic sol has been converted into one possessing typical lyophobic properties. Both alcohol and acetone are known to combine with water, and the explanation proposed for the results is that these organic substances are able to dehydrate the lyophilic particles. This will be accompanied, in the first place, by a decrease of viscosity. The stability of the "bare" particles will now be dependent on their electric charge, and so ions of opposite sign can cause flocculation. Another substance able to produce the same effects on an agar-agar sol is tannin; the amount required is now, however, quite small, e.g., about 1 per cent. It is suggested that the molecules of tannin are adsorbed at the interface of the dispersed particles, the glucose portion of the tannin being oriented toward the agar-agar, whereas the aromatic groups are toward the dispersion medium. As a result the external nature of the particle is completely changed, and there is a consequent decrease in hydration. Small amounts of alcohol, which are known to remove adsorbed substances (p. 1216), restore the original lyophilic properties to the agar-agar sol, although an excess once more brings about dehydration. These results seem to point to the conclusion that the stabil-

ity of a lyophilic sol in water is determined both by the electric charge and by hydration.

Salting Out.—The addition of large amounts of electrolytes to lyophilic sols results in the dispersed substance being precipitated; the effect is called "salting out," by analogy with the influence of electrolytes on true solutions (p. 729), and to indicate that the behavior is different from the ordinary coagulation generally observed with lyophobic systems. In the light of the foregoing interpretation of the stability of lyophilic sols, it may be suggested that in large concentrations electrolytes both dehydrate the particles and reduce their zeta-potential. It is generally accepted that the salting-out effect depends on the nature of the ions, and the salts of a given metal, for example, can be arranged in order of their decreasing ability to remove lyophilic substances from colloidal solution; this is sometimes called the Hofmeister series (p. 1259) or, more generally, the **lyotropic series**. The order of cations is

$\text{Mg}^{++} > \text{Ca}^{++} > \text{Sr}^{++} > \text{Ba}^{++} > \text{Li}^+ > \text{Na}^+ > \text{K}^+ > \text{Rb}^+ > \text{Cs}^+$,
and of anions,

$\text{Citrate}^{---} > \text{Tartrate}^{--} > \text{SO}_4^{--} > \text{C}_2\text{H}_3\text{O}_2^-$
 $> \text{Cl}^- > \text{NO}_3^- > \text{ClO}_3^- > \text{I}^- > \text{CNS}^-$.

This is very similar to the order of salting out of aniline, ethyl acetate and ether from aqueous solutions by various ions; an approximate valence effect is evident, but there is probably no connection between the salting out of lyophilic substances and the zeta-potential. As a general rule, the removal of the electrolyte causes the salted out material to return to colloidal solution, and it is probable that the cause of salting out of lyophilic substances is similar to that in true solutions (see p. 699).

Coacervation.—Salting out of a hydrophilic sol frequently gives a liquid aggregate, in place of a solid, and it has been known for some time that this often appears in the form of viscous drops, instead of a continuous liquid phase. It is now believed that this phenomenon, called **coacervation** (Latin: *heap together*) by H. G. B. de Jong and H. R. Kruyt (1929), is of common occurrence. The droplets, or **coacervates**, are formed most readily when two hydrophilic sols carrying opposite charges, e.g., gelatin at $\text{pH} < 4.7$ and gum acacia sols, are mixed in suitable amounts. It is believed that the shells of tightly bound water molecules surrounding the particles prevent them from coalescing, but the electrostatic attraction of their opposite charges holds a number of particles together in the form of a droplet, which may be microscopic or macroscopic in size. Re-dispersion of the coacervates can be brought about by the addition of iodide or thiocyanate ions which favor hydration of the particles, or by ions of high valence which diminish the zeta-potential of either sol. The suggestion has been made that coacervation phenomena are of importance in life process (H. G. B. de Jong, 1935).²¹

Protein Sols.—The sols of proteins are of special biological importance and their study presents some features of particular interest; gelatin will

be taken as typical, because it has been the chief subject of study. The isoelectric point of gelatin in the absence of appreciable amounts of electrolytes is at pH 4.7; in such a medium the particles are uncharged, and so there will be no electroviscous effect. The viscosity of gelatin sols has actually a distinct minimum at pH 4.7; any increase or decrease of pH causes a sharp increase of viscosity since the particles are then charged. In the presence of electrolytes, especially those of high valence, the position of the isoelectric point is changed; for example, at pH 4.4 the disperse phase of gelatin sol carries a positive charge, for reasons given above, but the addition of 0.42 millimole of potassium ferricyanide per liter results in the annulment of the zeta-potential and the system has, as before, a minimum viscosity although the pH remains at 4.4. Similarly, the isoelectric condition can be attained at pH 4.9 by addition of 0.07 millimole per liter of cobalt hexammine ion.

Addition of sufficient alcohol can bring about flocculation of a gelatin sol over a certain range of pH ; as is to be expected the minimum quantity is required at the isoelectric point, i.e., pH 4.7 in the absence of added ions. On either side of this pH the particles have a definite, if small, zeta-potential; hence a greater degree of dehydration, which is presumably the function of the alcohol, is necessary before flocculation can occur.

The protein casein provides an interesting illustration of relatively weak lyophilic behavior; whereas a gelatin sol is stable in the isoelectric condition, casein is not. The particles are apparently insufficiently hydrated to be stable in the absence of a zeta-potential. In solutions more acid or more alkaline than at the isoelectric point the casein sol is stable, but for a given potential much less alcohol will produce precipitation than is necessary for a gelatin sol. With the casein sol, therefore, stability depends on hydration and electrification simultaneously, for in the absence of one or other precipitation takes place readily. It may consequently be regarded as intermediate in behavior between lyophilic and lyophobic systems.

Intermediate Systems.—The behavior of casein sols approximates more closely to lyophilic systems, but others such as silicic and stannic acids, molybdenum and vanadium pentoxides, and the familiar Odén sulfur sols, prepared either by the action of hydrogen sulfide on sulfur dioxide in solution or by acids on sodium thiosulfate, are transitional cases resembling lyophobic sols. The valence rule for precipitation holds approximately, although marked exceptions are often evident, and there are distinct signs of the existence of a lyotropic series of ions of the same valence. As with true lyophilic systems the sensitivity to electrolytes is increased if alcohol or acetone is added. The sols are often more viscous than is to be expected for lyophobic systems, and when flocculation does occur it is frequently reversible, so that the solid is re-peptized when the precipitating electrolyte is washed away. It is possible that in these instances hydration contributes to some extent to the stability of the sol.

Size of Particles: The Ultracentrifuge.—As indicated on p. 1237 the size of colloidal particles, and particularly those of proteins, can be determined by a study of their distribution under the influence of an applied force. The method developed by T. Svedberg and his co-workers (1925 *et seq.*) is to rotate the solution in a specially designed high speed instrument called an **ultracentrifuge**. The force exerted on a particle in this manner can be as great as a million times gravity, or even more. If ω is the angular velocity of rotation, then the force exerted on a particle at a point distant x from the axis of rotation is $\omega^2 x/g$ times gravity. It follows, therefore, from (53) p. 498, that for a particle of radius r ,

$$\frac{dx}{dt} = \frac{2\omega^2 x r^2 (\rho - \rho')}{9\eta}, \quad (51)$$

where dx/dt is the velocity of the particle, at the point x , due to the action of the centrifugal force against the frictional resistance of the medium; ρ and ρ' are the densities * of the particle and the medium, respectively. Writing (51) in the form

$$\frac{dx}{x} = \frac{2\omega^2 r^2 (\rho - \rho')}{9\eta} dt, \quad (52)$$

and integrating, it follows that

$$\ln \frac{x_2}{x_1} = \frac{2\omega^2 r^2 (\rho - \rho')}{9\eta} (t_2 - t_1). \quad (53)$$

The quantity $(\ln x_2/x_1)/\omega^2(t_2 - t_1)$ is a constant for the given particle in the particular medium used; it is called the **sedimentation constant**, and given the symbol s , so that

$$s = \frac{2r^2(\rho - \rho')}{9\eta}. \quad (54)$$

The sedimentation constant is determined experimentally by passing a beam of light through the cell containing the system under investigation and allowing it to fall on a photographic plate. This is done after various intervals of time (t), without stopping the centrifuge, and the position (x) of the upper boundary of the particles is noted. If the value of s for successive intervals of time is constant then the system is "monodisperse"; that is to say, it contains particles of uniform size. For a "polydisperse" system it is often possible to detect two distinct boundaries, and from their positions at various times the respective sedimentation constants can be calculated. If s is known the radius of the particles can be derived from (54). The results may alternatively be stated in terms of the "molecular weight" (M) of the particle; if m is the actual mass then mN is equal to M , where N is the Avogadro number. Assuming the particle to be spherical, m is $\frac{4}{3}\pi r^3 \rho$, and so the molecular weight is $\frac{4}{3}\pi r^3 \rho N$. If this assumption is not justifiable, then the method described below must be employed to determine the particle weight.

The foregoing procedure is called the "sedimentation velocity" method. The size and molecular weight can be derived alternatively from an examination of the distribution of the particles when equilibrium is attained, that is, by the

* The quantities ρ and ρ' are strictly the reciprocals of the *partial* specific volumes of the particles (solute) and medium (solvent), respectively. For relatively large particles they may be taken as equal to the densities.

"sedimentation equilibrium" method. When the quantity of material driven outward by the centrifugal force is exactly balanced by the amount diffusing in the opposite direction, a state of equilibrium is reached and the position of the boundary of the particles will remain unchanged. If a particle moves a distance dx in time dt , then the amount of substance (dw) moving outward per sq. cm., under the influence of the centrifugal force, in the time dt is cdx , where c is the "concentration" of the particles; thus,

$$dw = cdx. \quad (55)$$

The diffusion coefficient (D) is the weight of material diffusing across a plane of 1 sq. cm. in area in unit time under a concentration gradient of unity (p. 260); hence the quantity dw diffusing across 1 sq. cm. in time dt when the concentration gradient is $-dc/dx$ is given by

$$dw = -D \frac{dc}{dx} dt, \quad (56)$$

which is the mathematical expression of Fick's law of diffusion. When sedimentation equilibrium is attained the quantities dw in (55) and (56) are equal but of opposite sign, so that

$$cdx = D \frac{dc}{dx} dt, \quad (57)$$

and hence

$$\frac{dc}{c} = \frac{1}{D} \frac{dx}{dt} dx. \quad (58)$$

By (45), p. 261, $D = RT/6\pi\eta rN$, and utilizing the value of dx/dt from (51), above, it follows that

$$\frac{dc}{c} = \frac{2}{3}\pi r^3 N \frac{\omega^2(\rho - \rho')}{RT} x dx \quad (59)$$

$$= \frac{M\omega^2(\rho - \rho')}{RT\rho} x dx, \quad (60)$$

where M is the "molecular weight" of the particle, and integration gives

$$\ln \frac{c_2}{c_1} = \frac{M\omega^2(\rho - \rho')(x_2^2 - x_1^2)}{2RT\rho}. \quad (61)$$

As in the sedimentation velocity method a beam of light is passed through the cell when equilibrium is attained, and if the absorption of light is assumed to be proportional to the concentration of the particles, the ratio c_2/c_1 can be estimated at two depths x_1 and x_2 , from the blackening of the photographic plate determined photometrically. Alternatively, a quantity proportional to $\ln c$, derived from the intensity of the plate at various points, may be plotted against x^2 and the slope of the straight line thus obtained gives $M\omega^2(\rho - \rho')/2RT\rho$, from which M may be calculated. If the system is not monodisperse the plot is not one line, but two or more straight lines, each corresponding to one set of particles; from the various slopes the molecular weights can be calculated. It is of interest to note that (61) is really a form of Perrin's equation (27), p. 257, for the equilibrium distribution of particles under the influence of gravity.

By means of the ultracentrifuge it has been found that many protein solu-

tions are monodisperse; molecular weights varying from about 18,000 to over 6,000,000 have been observed. It is probable that the particles in protein sols are single molecules and that these results represent the actual molecular weights.

The Svedberg form of ultracentrifuge is driven by an oil turbine, and the whole apparatus, including the means for photographing the solution without stopping the rotation, is very complicated. Attempts have, therefore, been made (e.g., J. W. McBain, *et al.*, 1935 *et seq.*) to utilize a simpler air-driven top as an ultracentrifuge. By the use of certain devices the liquid is immobilized so that when sedimentation equilibrium is attained the rotation can be stopped and the liquid at various distances from the axis of rotation analyzed. In this way it has been found possible to determine the molecular weight of substances in true solution, e.g., sucrose, with an error of only a few per cent. When the particles are relatively small, as in this instance, the correct value of ρ as $1/\bar{v}$, where \bar{v} is the partial specific volume of the solute, must be used.²²

THE PROPERTIES OF GELS

Gels.—The cooling of a lyophilic sol, e.g., of gelatin or agar-agar, provided it is not too dilute, results in the formation of a gel. Similarly a gel may be obtained by the addition of electrolytes, under suitable conditions, to lyophobic sols, e.g., silicic acid and ferric oxide, which exhibit some lyophilic character. In these systems the whole sets to an apparently homogeneous gel or jelly. It is probable that gels do not differ fundamentally in their properties and structure from gelatinous precipitates; in fact, whether one or the other is formed by the addition of electrolyte to a sol is often dependent on the experimental conditions (H. B. Weiser, 1924). The treatment here will be devoted particularly to homogeneous gels of the gelatin and silicic acid (silica gel) type. The process of gelation is obviously accompanied by a large increase of viscosity, but it is generally stated that the change is gradual and not a sudden one; in view of the uncertainty concerning the viscosity of lyophilic sols, and the complications introduced by the elasticity of the gel, this statement cannot be regarded as completely proved. There are, however, other indications that physical properties change gradually as a system passes from sol to gel form, and *vice versa*; for example, the curve of the refractive index against temperature for a gelatin sol (G. S. Walpole, 1913) shows no break when setting occurs, and no change is observed in the conductance of a soap solution when it sets to a gel, neither is there any change in the concentration of sodium ions nor of aqueous vapor pressure (M. E. Laing and J. W. McBain, 1920).

Two types of gel may be distinguished: they have been called **elastic** and **nonelastic**, or rigid, gels. Partial dehydration of an elastic gel, of which gelatin gel is a good example, leads to the formation of an elastic solid from which the original sol may be readily regenerated by the addition of water, and warming if necessary. Nonelastic gels, on the other hand, such as silica gel, become glassy, or fall to a powder, and lose their elasticity on drying; the sol cannot be obtained by mere addition of water

to the dry solid. It is perhaps in the phenomena of dehydration and re-hydration of the partially dried gels, that the most important distinction between elastic and nonelastic gels lies. If a gelatin gel is dehydrated the vapor pressure decreases continuously, and re-hydration by exposure to water vapor is approximately reversible over the whole range from completely dry to wet gel. A silica gel, on the other hand, shows discontinuities, and re-hydration is reversible only for the almost dry gel. The phenomena concerned have been the subject of considerable study by J. W. van Bemmelen (1880 *et seq.*) and by R. Zsigmondy (1911-14), and the results are of interest, although outside the scope of this book. The main difference between elastic and nonelastic gels is attributed to the rigidity of the walls of the capillaries formed when the latter are dehydrated, whereas with the former they are supple. Dehydrated silica gel is so honeycombed with fine capillaries that it is a most valuable adsorbing agent which has found application both in industry and in the laboratory.²³

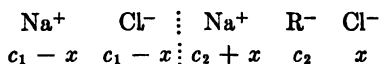
Imbibition.—When placed in a suitable liquid a gel is able to imbibe considerable volumes of the liquid, often several times as large as the volume of the original gel. The imbibition of the liquid is thus accompanied by a large increase of volume, so that the phenomenon is often called **swelling**. It may result in the production of considerable pressure; e.g., if dry gelatin is placed in a porous pot and stood in water the swelling will break the pot. In spite of the increase in volume of the gel, the net volume of gel and water decreases in the imbibition process. The influence of salts on swelling was studied by F. Hofmeister (1891), who found the influence of various anions to correspond to the familiar lyotropic or Hofmeister series (p. 1254). The anions sulfate, tartrate, citrate and acetate inhibit the swelling of gelatin and similar gels, the inhibiting effect decreasing in the order given. On the other hand, chloride, chlorate, nitrate, bromide, iodide and thiocyanate favor the imbibition of water, the effect of the chloride ion being least. When swelling is considerable, as in solutions of iodides, the gel often disperses at room temperatures, and so forms the sol spontaneously; in other cases the sol is formed by warming the swollen gel. The Hofmeister series thus gives the order of the temperatures to which the gel must be heated, in the presence of various anions, before it changes into a sol. It is for this reason that the term “lyotropic” (Greek: *change to liquid*) is used for this series of anions.

The extent of swelling varies with the *pH* of the medium and is a minimum at the isoelectric point; J. Loeb (1920-22) has suggested an explanation of these facts in terms of the Donnan equilibrium.

The Donnan Membrane Equilibrium.—If a solution of an electrolyte consisting of two diffusible ions is separated by a membrane from another solution containing a salt with a nondiffusible ion, e.g., a protein ion, then at equilibrium the distribution of the former will be unequal on the two sides of the membrane. The thermodynamic proof of this statement was first given by

F. G. Donnan (1911), who later verified it experimentally. Suppose a solution of sodium chloride, concentration c_1 , is separated by a diffusion membrane from a solution of a salt NaR , concentration c_2 , R^- being a nondiffusible ion. At equilibrium a certain amount (x) of sodium and chloride ions will have passed through the membrane, so that the final state may be represented as follows:

Equilibrium State



It should be noted that the condition of electrical neutrality applies on both sides of the membrane, so that the total concentration of positive ions is equal to that of the negative ions in each solution. At equilibrium the chemical potentials of the substance which is present on both sides of the membrane, i.e., sodium chloride, must be equal, i.e., $\mu_{\text{NaCl}(1)} = \mu_{\text{NaCl}(2)}$, where the subscripts (1) and (2) refer to the two solutions. As seen on p. 962 the chemical potential of an electrolyte may be taken as the sum of the potentials of its ions; hence,

$$\mu_{\text{Na}^+}^0 + RT \ln a_{\text{Na}^+(1)} + \mu_{\text{Cl}^-}^0 + RT \ln a_{\text{Cl}^-(1)} = \mu_{\text{Na}^+}^0 + RT \ln a_{\text{Na}^+(2)} + \mu_{\text{Cl}^-}^0 + RT \ln a_{\text{Cl}^-(2)}; \quad (62)$$

$$\therefore a_{\text{Na}^+(1)} \times a_{\text{Cl}^-(1)} = a_{\text{Na}^+(2)} \times a_{\text{Cl}^-(2)}, \quad (63)$$

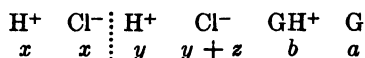
where the a terms represent the activities of the indicated ions in the two solutions. If these are dilute, the activities may be replaced by concentrations, so that

$$\begin{aligned} c_{\text{Na}^+(1)} \times c_{\text{Cl}^-(1)} &= c_{\text{Na}^+(2)} \times c_{\text{Cl}^-(2)}; \\ \therefore (c_1 - x)(c_1 - x) &= (c_2 + x)x; \end{aligned} \quad (64)$$

$$\therefore \frac{x}{c_1} = \frac{c_1}{c_2 + 2c_1}. \quad (65)$$

The fraction x/c_1 gives the proportion of sodium chloride initially present which has diffused through, and this is seen to be smaller the larger the concentration c_2 of the nondiffusible ion.

From the point of view of lyophilic colloids it is of interest to consider a solution containing the hydrochloride of a protein, e.g., gelatin, separated by a membrane from a solution of hydrochloric acid; the distribution at equilibrium will then be as shown below.



The nonionized or dual-ion form of the amphoteric gelatin is represented by G , and GH^+ is the corresponding cation. If the protein has n basic groups, then b is equal to z/n , where z is the concentration of chloride ions and b that of GH^+ ions produced by the gelatin hydrochloride. Using the simple Donnan equation (64), applicable to dilute solutions, it follows that

$$x \times x = y(y + z); \quad (66)$$

$$\therefore z = (x^2 - y^2)/y. \quad (67)$$

The total concentration of particles, i.e., ions and molecules, on the right of the membrane is $2y + z + a + b$ and on the left it is $2x$, so that if the membrane

formed part of an osmometer, there would be a difference of osmotic pressure on the two sides determined by the concentration $2y + z + a + b - 2x$. If the molecular weight of the protein is high and it has a number of basic groups, a and b can be neglected, and assuming the solutions are dilute enough for the van't Hoff equation (p. 671) to be used, the osmotic pressure Π is given by

$$\Pi = RT(2y + z - 2x). \quad (68)$$

This can be written in two alternative ways: one in which z is eliminated by means of (67), and another in which y is replaced by its equivalent form derived from (66); thus,

$$\Pi = RT(x - y)^2/y = RT(\sqrt{4x^2 + z^2} - 2x). \quad (69)$$

In the presence of concentrated acid, i.e., when x is large, $\sqrt{4x^2 + z^2}$ is not much greater than $2x$, and hence the osmotic pressure set up between the acid solution and that of gelatin will be small. As x decreases, however, the osmotic pressure will increase, but it must reach a maximum and decrease again almost, to zero when y becomes equal to x . These conclusions reached from the Donnan theory of membrane equilibrium have been verified by J. Loeb in his comprehensive studies of the influence of electrolytes on the properties of gelatin. The latter was separated from an acid solution by a collodion membrane, and the excess pressure attained at equilibrium for various hydrogen ion concentrations was determined. In agreement with theory, the osmotic pressure on the protein side of the membrane was found to increase to a maximum and then to decrease as the pH of the medium was increased, i.e., as the acidity was diminished. It will be seen from (67) that when x is equal to y , that is, when the osmotic pressure sinks almost to zero, z should be zero; in other words, this condition arises when the protein is virtually nonionized. According to the statement on p. 998, the ionization of an amphoteric electrolyte as acid and base is a minimum at its isoelectric point; hence the value of z is then a minimum and so also will be the osmotic pressure. This anticipation has been verified by experiment. At the isoelectric point of the protein the distribution of acid is the same on both sides of the membrane, i.e., $x = y$, and the osmotic pressure is that due to the dissolved protein alone, i.e., to the concentration a which was neglected in deriving (68). A similar condition can be realized by adding a relatively large amount of diffusible electrolyte to the system, because the Donnan equilibrium then requires x and y to be almost identical. Consequently, in the determination of the molecular weight of a protein by the osmotic pressure method it is necessary either to work at the isoelectric point, or in the presence of excess of an electrolyte which diffuses through the membrane (cf. p. 673).

The possibility that membrane equilibrium might be an important factor in the swelling of gels was first considered by H. R. Procter and J. A. Wilson (1914, 1916), and developed by J. Loeb (1920-22). The essential point of the argument is that when placed in water the gel, assumed to be an amphoteric substance, unless it is at its isoelectric point, will set up a Donnan equilibrium, the gel itself acting as the membrane. From the discussion above it is clear that the total ionic concentration will be greater within the gel than in the surrounding medium, and hence osmosis will occur and water will be drawn into the gel and so cause swelling. The extent of osmosis varies with the pH of the external solution, and so the swelling should vary in an analogous manner,

increasing to a maximum and sinking again at the isoelectric point; this has been found to be generally true in practice. According to Loeb the influence of the Hofmeister series of ions on various properties of protein sols and gels, e.g., precipitation, osmotic pressure and swelling, consists of two parts. In the first place they affect the pH of the medium, and if the hydrogen ion concentration is adjusted to be the same in each case, the various anions fall into simple valence groups. A similar result is claimed for cations, the effect of any ion being determined by its valence, at constant pH, and not by its nature. In the second place there is a specific effect of different ions of the same valence, just as is found in the properties of true solutions; this factor may, however, be of secondary importance.

If Loeb's conclusions were correct, the Hofmeister series would have no significance, but it now seems to be generally agreed that the Donnan equilibrium is responsible for only a portion of the swelling. This is evident from the observation that even at its isoelectric point, gelatin exhibits considerable imbibition, although the extent is small compared with that at other pH values. Nevertheless, the swelling is far from zero, as required by the Procter-Wilson-Loeb theory. The Loeb valence rule concerning the influence of ions is probably applicable to that part of the swelling due to the osmotic pressure difference, but this does not include other factors responsible for imbibition. The fact that rubber swells considerably in benzene, where ionic equilibria are not operative, shows clearly that imbibition cannot be of a purely osmotic character.²⁴

Structure of Gels.—Various theories of the structure of gels have been proposed since C. von Nägeli (1858) first attempted to account for their properties. The observations of W. Bachmann (1911–12) of the changes, seen in the ultramicroscope, in gelatin sols during cooling have shown that in the course of gelation the colloidal particles increase in size, while at the same time the translational motion ceases and the Brownian movement becomes restricted. When the sol has set to a gel the relatively large particles appear to cohere and are quite stationary. In some cases, e.g., agar and probably gelatin, the formation of the gel is accompanied by an increase of hydration of the particles. This is indicated by the fact that substances such as alcohol and soluble salts, which affect hydration (cf. p. 1253), have a marked influence on the conditions of gelation. Further, if an agar-agar sol is cooled to below 40° c. for a short time, the viscosity is found to have suffered an increase when the temperature is raised once again, suggesting a more or less permanent increase of solvation upon cooling. One view is that in the course of gelation the particles in the sol unite to form a number of chains or threads (fibrils), which become interlocked so that the viscosity of the system increases and eventually a semi-solid form is acquired. Part of the dispersion medium may be involved in solvation, but the major portion is believed to be held by capillary forces between the fibrils. Since the rate of diffusion of ordinary solutes is almost the same in pure liquid as in a dilute gel in the same medium, it is evident that the gels must contain "pores" of solvent; it has been estimated from vapor pressure measurements that these have diameters of the order of 5 μ , and so the

liquid might easily be retained by capillary effects. A modification of the fibrillar theory is that the gel consists of a three-dimensional network, formed by linking and cross-linking of molecular chains. The more or less sudden gelation upon cooling the sol is then analogous to the rapid branching of chains, accompanied by explosion, in certain gas reactions (p. 1083). Support for this view is provided by observations on high polymers, consisting of a polyhydroxy-compound, e.g., glycol or glycerol, and a dibasic acid. The molecules have three or more "functional" groups, and hence form three-dimensional polymers; these become gels presumably when the extent of branching or cross-linking exceeds a certain critical value. Polymers having two functional groups, on the other hand, do not yield gels, since they can form linear structures only. The foregoing suggestions account for the elasticity of gels, and also for the gradual transition in physical properties from sol to gel.

All types of gels appear to have analogous structures; whether a gel is rigid or elastic when dry depends on the physical and chemical nature of the disperse phase. In some cases, e.g., gels of sparingly soluble salts, such as strontium sulfate in aqueous alcohol (P. P. von Weimarn, 1928), the network undoubtedly consists of material of a crystalline character, but in others, e.g., gelatin or agar, the molecules are arranged in a more random manner. X-ray examination of silica gel has shown it to consist of a mass of small crystallites (see p. 518).²⁵

Syneresis.—Many gels exhibit the phenomenon, first observed by T. Graham (1861) and called by him *syneresis*, of exuding small amounts of liquid on standing; at the same time the gel contracts. Syneresis is shown by concentrated silicic acid gels and by dilute gels of gelatin and agar-agar. It appears that with certain gels the phenomenon becomes evident with increasing concentration whereas with others dilution is necessary. There is no net volume change as a result of syneresis, and so it should not be described as the reverse of imbibition, as has been suggested. The mechanism of the process is not clearly understood, but syneresis is believed to be due to the exudation of water held by capillary forces between the heavily hydrated particles constituting the framework of the gel.

Thixotropy.—When suitable small quantities of electrolyte are added to a concentrated ferric oxide sol, a pasty gel is formed which has the remarkable property of liquefying when shaken, only to set again to the gel form on standing (A. Szegevari and E. Schalek, 1923). This phenomenon, subsequently observed in other colloidal systems, e.g., alumina, vanadium pentoxide, zirconium dioxide, stannic oxide, and certain gelatin sols, and even with suspensions of fine clays, particularly *bentonites*, has been called *thixotropy* (Greek: *change by touch*) by T. Péterfi (1927). A thixotropic gel can be liquefied by ultrasonic, i.e., high frequency sound, waves, and in fact by any means which produces internal mechanical stress. As may be expected the phenomenon of thixotropy is dependent on suitable concentration of electrolyte; if the concentration is too high

the sol is precipitated in a flocculent form which cannot be dispersed by shaking.

Several theories have been proposed to account for thixotropy. According to the **solvation theory** thick envelopes, called "lyospheres," of oriented water molecules are formed round the colloidal particles and these eventually become so large that freedom of movement is lost and the system becomes a gel. It is postulated that the lyospheres are partially destroyed on shaking so that the gel liquefies. The fact that suspensions of finely divided inorganic substances in organic media exhibit thixotropy is, however, a strong argument against the solvation theory. Another point of view is the **oriented coagulation theory**; the particles of sol capable of forming a thixotropic system are believed to be anisotropic or anisometric, or both, so that the electric charge and water of hydration, if any, are unequally distributed. When the correct amount of electrolyte is added to reduce the zeta-potential to a sufficient extent, the particles tend to coalesce, but owing to their shape they can form a stable gel only if correctly oriented. The extent of the surfaces of the particles in contact is probably limited as a consequence, and any stress will tend to reduce this and hence destroy the gel structure. It is of interest in connection with this theory to record that nearly all thixotropic sols show marked streaming double refraction,* so that the particles are probably anisotropic and anisometric. A further observation indicating the importance of the shape of the constituent particles is that small flat or rod-like particles, e.g., of kaolin, graphite, etc., readily form thixotropic systems, but approximately isometric particles of quartz, feldspar and fluorspar do not. A third theory is based on the hypothesis of three-dimensional gel structure; it is supposed that when a gel is thixotropic, the shaking is sufficient to break the cross-linkings, thus re-forming the sol. The particles would then be essentially linear, thus accounting for the double refraction of flow.²⁶

Precipitation in Gels: Liesegang Rings.—It was observed by R. E. Liesegang (1896) that when a drop of concentrated silver nitrate solution was placed on a film of gelatin containing potassium dichromate, the silver chromate was not precipitated continuously but in the form of a number of concentric rings; rhythmic precipitates of this kind are called **Liesegang rings**. Much work has been done on this subject and the ring formation has been observed with many precipitates and with gels of different materials. The phenomena are best studied by dissolving one of the reactants, e.g., potassium dichromate, in a warm gelatin sol which is then poured into a test tube and allowed to set. The other reactant, e.g., silver nitrate, is next placed on top of the gel; the silver chromate precipitate is then seen to form as a series of bands with spaces of clear gel between. Liesegang rings have also been obtained in gels of agar-agar and of silicic acid; in the latter case the reactant is dissolved in the mineral acid solution before addition of the sodium silicate. Besides silver chro-

* When a liquid containing rod-like particles flows through a narrow tube, there is a tendency for the particles to align themselves with their long axes parallel to the direction of flow. This results in a difference of refractive index in two directions at right angles; this is "streaming double refraction" or "double refraction of flow" (cf. p. 346).

mate, rhythmic precipitates of numerous other substances have been prepared; examples are magnesium hydroxide, lead iodide, lead chromate, mercuric iodide, copper chromate and copper ferrocyanide. Liesegang rings of gold are formed when gold chloride dissolved in a silica gel is reduced by oxalic acid in the presence of light; the purplish precipitate does not form in the dark.

The nature of the gel is not of fundamental importance, since banded structures have been obtained by precipitation of many salts in aqueous solution in the absence of any gel-forming substance, provided there is a restriction to movement by convection, etc., as in a capillary tube. Nevertheless, the presence of the gel appears to facilitate rhythmic, in preference to continuous, precipitation, and the influence is to some extent specific. For example, silver chromate readily shows the Liesegang phenomenon in gelatin gel, but not in agar-agar; with lead chromate the reverse is true; further, neither forms bands readily in silica gel. Similarly, the presence of impurities in the gel, e.g., acid in gelatin, often has a marked effect on the ease of band formation. By suitable adjustment of conditions, e.g., concentration of electrolyte, it appears that the specific influence of the gel material may often be nullified. Banded effects are frequently found in nature, and it is possible that rhythmic precipitation may be a factor of importance in this connection. Liesegang developed the view that the banded agates were formed in a manner analogous to the rhythmic precipitates in gelatin, but this interpretation is not universally accepted. It is important to remember that periodic structures are by no means uncommon, and they are not necessarily all due to the same cause.

Various theories have been proposed from time to time to account for the Liesegang phenomena. None of these alone appears to be capable of explaining all the observations and it is most probable that several different influences are operative. Although the presence of the gel is not essential, it is quite definitely an important contributory factor in the formation of banded precipitates and so a rôle must be attributed to it. There are three directions in which the gel may influence precipitation: first, colloidal material is known to retard crystallization in general (Wilhelm Ostwald, 1897); second, in the presence of the colloid finely divided precipitates are formed which will adsorb ions of the reactant (S. C. Bradford, 1916); and third, the semi-rigid nature of the gel assists in the maintenance of concentration gradients (H. N. Holmes, 1918). All these factors would account for the specific influence of the gel material, but supersaturation, concentration gradients, and adsorption of ions by the precipitate could all occur in the absence of gel; it is thus possible to account for the part played by the gel, without admitting its presence to be essential.

On the basis of the factors mentioned, a simplified explanation of the Liesegang phenomena, e.g., for the precipitation of silver chromate, would be some-

what as follows. As the silver ions diffuse into the chromate contained in the gel, reaction occurs but the silver chromate remains for a time in the colloidal state and a visible precipitate does not form. The depletion of the chromate concentration causes chromate ions to diffuse rapidly in the vicinity of the reaction zone, leading to a region of relatively low concentration. When precipitation of silver chromate occurs, perhaps because of the accumulation of the ions of potassium nitrate, the other reaction product, some of the potassium chromate may be adsorbed leading to a further depletion of concentration. Since the rate of diffusion is most rapid where the concentration gradient is greatest, there will be a region in the vicinity of the precipitation zone where the concentration of chromate ions is much less than at a further distance. It will, therefore, be necessary for the silver ions to diffuse some distance into the gel before reaching a zone in which the chromate concentration is once more sufficient for the colloidal silver chromate to be formed and subsequently precipitated. It was originally suggested by Wilhelm Ostwald (1897) that prior to precipitation the silver chromate existed as a supersaturated true solution, but E. Hatschek (1912) showed that Liesegang rings could be obtained in gels already containing crystals of the precipitated substance, so that ordinary supersaturation was impossible. The postulate that the substance is present in a colloidal state before a visible precipitate is formed, however, overcomes this difficulty. The explanation given here probably presents an over-simplified picture of the Liesegang phenomenon, for many observations appear to be difficult to interpret in this way.²⁷

COLLOIDAL ELECTROLYTES

Soaps.—The soaps, that is, the sodium and potassium salts of long-chain fatty acids, exhibit a remarkable behavior in aqueous solution,* which has earned for them the title of **colloidal electrolytes** (J. W. McBain, 1920); similar properties occur in other compounds having a long hydrocarbon chain (R) terminating in a group capable of ionization, e.g., the alkyl sulfonic acids (RSO_3H), alkyl sulfates (ROSO_3K), and quaternary ammonium salts (RNMe_3Br), and in certain dyestuffs, e.g., congo red, which contain an ionizable group attached to a large organic residue. The earliest measurements of the rise of boiling point of aqueous soap solutions (F. Krafft, 1894) and of the lowering of the vapor pressure (A. Smits, 1902–03) indicated such small osmotic effects as to suggest the whole of the material was in colloidal solution. It was shown, however, by J. W. McBain who, with the aid of a number of collaborators, has carried out since 1911 a series of comprehensive investigations on the properties of soap solutions, that the earlier measurements of boiling point and vapor pressure were in error because of the presence of air in the solutions. By means of the dew-point method (p. 632), reliable data concerning osmotic effects of soap solutions have been obtained (J. W. McBain and C. S. Salmon, 1920); from the results the corresponding elevations of the boiling point at 90°C . have been calculated, and are given in Table 205. The osmotic effect of a soap solution, at a concentration of about 0.5 N, is seen to be somewhat less than that of a nonelectrolyte.

* In alcoholic solution soaps generally behave as simple electrolytes.

In spite of this similarity, the solutions of the sodium and potassium salts of the higher fatty acids have a very marked electrical conductance. The older measurements of L. Kahlenberg (1898) were at one time believed to be incorrect because of hydrolysis of the soap and the presence of impurities, but J. W. McBain (1911 *et seq.*) showed by hydrogen electrode and catalytic measurements (p. 989) that hydrolysis was in fact almost negligible, and that the alkali metal salts of stearic and palmitic acids had a conductivity of the same order as that of the salts of the simpler fatty acids. Some of the results for potassium salts are recorded in Table 205, together with the osmotic effects mentioned above; data for

TABLE 205. OSMOTIC PROPERTIES AND CONDUCTANCES OF 0.5 N SOLUTIONS OF POTASSIUM SALTS AT 90° C.

Salt	Number of Carbon Atoms	Rise of Boiling Point	Equivalent Conductance
Stearate	18	0.17° C.	113 ohms ⁻¹ cm. ²
Palmitate	16	0.19°	127
Myristate	14	0.23°	135
Laurate	12	0.26°	145
Acetate	2	0.46°	196
Nonelectrolyte	—	0.24°	—

potassium acetate and a nonelectrolyte are included for comparison. It is clear from these figures that although soap solutions have the osmotic properties of a nonelectrolyte, they nevertheless possess a conducting power of the same order as the acetates and other simple salts. This abnormal behavior can be detected with the salt of the fatty acid containing six carbon atoms and, as seen from the table, becomes more apparent with increasing chain length.

Ionic Micelles.—To explain these and other results, J. W. McBain (1913 *et seq.*) developed the idea that in very dilute aqueous solutions the soaps behave like ordinary salts and are considerably ionized into an alkali metal cation and a fatty acid anion. At appreciable concentrations, however, the anions are believed to aggregate together to form **ionic micelles**, containing ten or more ions together with an appreciable number of water molecules. In view of what has been said concerning the properties of long-chain compounds (pp. 485, 1211) the formation of an aggregate with the hydrocarbon chain (R) inward and the polar or ionic portion ($-\text{COO}^-$) outward, toward the dispersion medium, is readily understood. Further, the relatively high viscosity of soap solutions suggests that the particles are extensively hydrated. Since the ionic micelles have large molecular weights, their osmotic contribution will be negligible and the observed effects will be due almost entirely to the alkali metal ion; the rise of boiling point, etc., should thus be, as found experimentally, similar to that for a nonelectrolyte. In spite of their large size the ionic

micelles would be expected to have a high conductance, for the following reason. If n spherical ions of radius r aggregate to form a closely-packed spherical ionic micelle, the radius of the latter will be $rn^{1/3}$. According to Stokes's law the resistance to motion of a spherical particle is proportional to its radius; hence the ratio of the resistance to motion of the ionic micelle to the total for the n individual ions which constitute it is $rn^{1/3}/n$, i.e., $n^{-2/3}$. Since the micelle carries the same total charge as the separate ions, it is clear that its conducting power, which is inversely proportional to the resistance to motion, will be increased $n^{2/3}$ times as a result of aggregation. This calculation is, of course, approximate for it involves the assumption of spherical ions, close packing and no volume change due to hydration, but it shows that the ionic micelle may have a large molecular weight and yet have a conductivity even greater than that of all the individuals of which it is composed.

By subjecting soap solutions to ultrafiltration under pressure, using hardened collodion membranes of differing pore size, J. W. McBain and W. J. Jenkins (1922) were able to prove the presence of two distinct types of colloidal particles. These results have been confirmed by means of X-ray measurements. At moderate concentrations the soap is largely present in the form of approximately spherical particles, corresponding to the ionic micelles described above, with a diameter of about 40 to 50 \AA . As the concentration is increased, there is an increasing formation of larger particles, similar to those found in ordinary colloidal solutions. These particles are crystalline in nature having, as might be expected, a laminated structure (cf. p. 515) of layers of undissociated molecules of soap. The term **neutral colloid** was originally proposed, but the expression "lamellar colloid" has also been suggested, because of the structure of the colloidal particles. In the neutral colloid form the conductance would be very small, being of the same order as that of a normal colloidal system. It must be understood, of course, that even the "neutral" colloid is charged, as a result of the attachment of ions, just as is any other colloidal particle.

The foregoing views concerning the structure of colloidal electrolytes have been amplified and confirmed by electrical conductance measurements. From an examination of sulfates, sulfonic acids and alkyl ammonium bromides with long paraffin chains, it appears that the variation of equivalent conductance with concentration is, in general, to be represented by a curve of the type shown in Fig. 269; with ordinary soap solutions the initial, relatively flat, portion is not easily detected because of hydrolysis. There is little doubt that in the range AB , that is, up to a concentration of about 0.001 to 0.01 N , the system behaves essentially as a simple uni-univalent crystalloidal electrolyte; at B , however, where the curve shows a marked bend, ionic micelle formation commences and rapidly becomes the predominating factor. If the conductance is measured with a current of high voltage, viz., 200,000 volts per cm., the curve rises at B , as shown by the broken line, instead of falling (J. Malsch and G. S. Hartley, 1934); this result suggests that the highly conducting ionic micelles are formed in the vicinity of B . It is known (p. 908) that the use of a high voltage in the

measurement of conductance eliminates the influence of the ion atmosphere, and the fact that the equivalent conductance then rises above the value at infinite dilution shows that the simple ions present in the solution are being replaced to a great extent by a species, presumably ionic micelles, with a greater conducting power. This view is supported by transference number determinations which show that the contribution of the long-chain ions increases appreciably at concentrations in the vicinity of B .

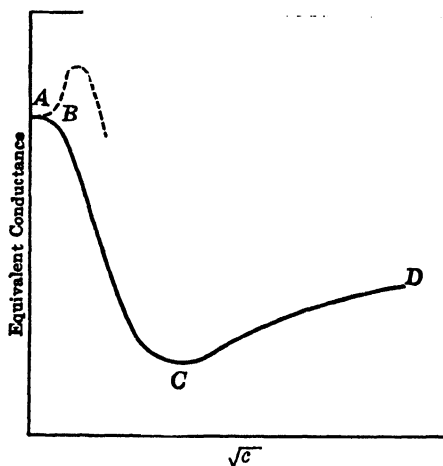


FIG. 269. Conductance of colloidal electrolyte

The fall in equivalent conductance from B to C , with low-voltage current, must be partly due to the effect of the ion atmosphere, but since the conductance decreases rapidly, after passing through a maximum, even when high voltages are used, another factor must be operative. This is presumably the formation of the poorly conducting neutral colloid as the concentration of the colloidal electrolyte is increased. The remarkable increase in conductance from C onward, for concentrations greater than about 0.05 to 0.1 N , is more difficult to explain. Some writers attribute it to the production of increasing amounts of ionic micelles from the neutral colloid, while others have suggested that it is due to the liberation of some of the attached "gegen-ions" or "counter-ions." However, it must be admitted that neither interpretation can be regarded as complete.

Colloidal Properties.—The presence of the various particles in soap solutions accounts for the undoubtedly colloidal properties. Although in common with certain other sols they do not show any appreciable heterogeneity in the ultramicroscope, relatively concentrated solutions readily set to gels on cooling, and particles of different sizes may be separated by ultrafiltration. Addition of considerable amounts of electrolytes produces the well-known salting out of soap, which is employed in the manufacture of that commodity. In addition to the reversible gelation as a result of temperature changes, soap solutions have another property in

common with hydrophilic sols, namely a high viscosity; this fact, like the ready formation of gels, indicates considerable hydration of the micelles. The question of the structure of soap gel need not be discussed here, as these substances are typical gels.²⁸

COLLOIDAL SYSTEMS IN ORGANIC MEDIA

Organosols.—Relatively little work has been done on colloidal systems in organic dispersion media, frequently called **organosols**; this is because of the inherent limitations of the systems and the difficulties associated with their study. T. Svedberg (1906 *et seq.*) has used the electrical dispersion method (p. 1274) to obtain sols of a number of metals in various organic liquids, e.g., ether, pentane, chloroform and isobutyl alcohol, and other procedures have also been used to prepare organosols of metals. These systems are lyophobic in behavior and are relatively unstable; precipitation often takes place on warming or on adding another organic liquid. The particles can be detected in the ultramicroscope, and are seen to display the Brownian movement. Organosols show only slight or indefinite electrophoresis and it is unlikely that metal sols in hydrocarbon media are electrically charged, for it is not easy to see how such a charge could originate. This raises the question of the factors governing the stability of organosols. The view has been expressed (J. K. Giles and C. S. Salmon, 1923), with some justification, that organosols can only be formed in the presence of an impurity which acts as a stabilizer. If the dispersion medium is of a polar type, the formation of ions is not impossible and the organosols may depend for their existence on the same factors as are applicable to hydrosols. It is of interest that T. Svedberg (1907) found that a platinum sol in pure ether coagulates almost immediately, but the presence of small quantities of water confers stability on the system. Further, the technically important colloidal solution of graphite in a hydrocarbon oil cannot be obtained by direct peptization; the graphite is first dispersed in water, with the aid of tannin as protective agent, and the water is subsequently displaced by oil.

Where ionization is not possible, e.g., in a hydrocarbon medium, the presence of a minute amount of a substance which is adsorbed by the colloidal particles and lowers the interfacial tension might serve to prevent coagulation; the effect would be similar to the influence of soap and similar compounds in stabilizing emulsions (p. 1276). This factor may also be operative in polar media, as for example in the silver sol obtained by reduction with hydrogen of silver stearate dissolved in pure stearic acid. Sodium sulfate has no precipitating effect, and so it is unlikely, in spite of the possibility, that ions are responsible for the stability of the sol. A number of colloidal solutions of metals in nonaqueous solvents have been prepared by reduction of solutions of metallic salts by means of phosphorus, hydrazine, stannous chloride, tannin, etc. (cf. p. 1273); in alcohol, acetone or dioxane a small quantity of cellulose acetate acts as

the stabilizing agent, while in benzene and ether the stabilizer is rubber. The sol probably owes its existence to this added lyophilic colloidal material.

A relatively large number of lyophilic organosols involving organic compounds of high molecular weight are known, and many of them, e.g., sols of cellulose esters, are of industrial importance. Cellulose nitrate and acetate readily imbibe a variety of organic polar solvents and on warming colloidal solutions are formed; when these are cooled they frequently set to gels. Similarly, sols and gels can be obtained of rubber and of aluminum and magnesium soaps in hydrocarbons and other non-polar solvents. It is known that high molecular weight substances of a hydrocarbon nature, e.g., rubber, or those consisting mainly of a hydrocarbon group, e.g., soaps of polyvalent metals, imbibe and are peptized by nonpolar liquids, whereas the cellulose esters require media such as alkyl and other esters, acetone and nitro-compounds, which contain polar groups. Since a given compound is peptized only by a liquid having a related structure, it is probable that the stability of lyophilic organosols is to be ascribed entirely to solvation, electrical charge having relatively little, if any, influence; these sols are thus similar to hydrophilic colloidal solutions at their isoelectric points. The fact that organosols can be precipitated by the addition of a liquid which dissolves in the dispersion medium but does not dissolve the disperse phase, provides some support for the view expressed. For example, methyl or ethyl alcohol, or acetone, brings about precipitation of rubber from a hydrocarbon sol. The added liquid presumably functions as precipitant by removing some of the molecules of solvation of the dispersed rubber particles. Cellulose acetate sols behave in an exactly similar manner. There is an additional possibility which must be borne in mind: substances which form lyophilic organosols have very high molecular weights, and it is possible that the particles are single molecules, as in a true solution. The liquids causing peptization are just those which might be expected to act as solvents, while those which would be expected to decrease the normal solubility bring about precipitation.

It is an interesting fact that whereas neither ether nor alcohol alone will peptize pyroxylin (nitrated cellulose) at ordinary temperatures, a mixture of the two liquids is effective. In explanation of this fact it has been suggested (A. Highfield, 1926) that since the nitrated cellulose contains both polar, e.g., $-\text{OH}$, $-\text{ONO}$, and nonpolar portions the best solvent is one also possessing both polar and nonpolar radicals. The former is provided by the alcohol and the latter by the ether. A mixture of alcohol and benzene, however, does not peptize pyroxylin, so presumably the benzene is too nonpolar. The interpretation of the efficacy of the alcohol-ether mixture may not be exact, but it certainly provides a rough guide to the actual behavior and accounts for the observation that acetone containing 9 per cent of water is a better peptizing medium for cellulose nitrate (12.2 per cent nitrogen) than is acetone alone; water in

excess of this amount makes the medium too polar and peptization is decreased. It is evident that, in a general way, swelling and subsequent peptization of a solid will occur when the adhesional energy between the solid and liquid is greater than the cohesional energy of the solid at a large number of points. It is known that adhesional energy is greatest between two substances possessing similar groups (p. 485), polar or non-polar, and so the peptizing effects described above are readily understood.²⁹

PREPARATION OF COLLOIDAL SOLUTIONS

Many substances, particularly those of high molecular weight forming lyophilic sols, go into colloidal solution when they are warmed with a suitable dispersion medium. Lyophobic sols, on the other hand, have generally to be prepared by special methods yielding particles of the appropriate size; various procedures have been used which may be conveniently considered under the headings of condensation methods and dispersion methods.

Condensation Methods.—The materials from which the sol is prepared are originally present in true solution as ions or molecules, and as a result of chemical reaction between them particles of colloidal size are obtained. The types of chemical reaction and general principles which have been employed will be described briefly. The factors determining the "condensation" of molecules and ions into particles of appreciable size have been examined by P. P. von Weimarn (1908 *et seq.*) who suggested the relationship

$$\text{Initial rate of condensation } (V) = k \frac{\text{Condensation pressure}}{\text{Condensation resistance}},$$

where k is a constant. The normal solubility (S) of the substance separating out from solution may be taken as a measure of the condensation resistance, whereas the condensation pressure may be assumed to be equal to the difference between the total concentration (Q) momentarily produced as a result of chemical reaction and the normal solubility (S); that is,

$$V = k \frac{Q - S}{S}. \quad (70)$$

If the conditions are such that V is large, that is, the rate of condensation is high, the substance will separate from solution in the form of very small particles; if V is of the order of 100,000 to 200,000, with k equal to unity, the particles separating are of colloidal dimensions. There are clearly two ways whereby the rate of condensation may be increased: either by making Q large or S small. To make Q large the solution, or solutions, from which the particles separate should be very concentrated. For example, if equal volumes of 7 N barium thiocyanate and manganese sulfate are mixed, the momentary concentration (Q) of barium sulfate will be 3.5 N , whereas the normal solubility (S) is 2×10^{-5} N . Under these conditions V is 175,000 by (70), with k equal to unity, and the particles of barium sulfate are undoubtedly of colloidal dimensions, but since the concentration of electrolyte is very high, immediate precipitation occurs and a gel is obtained. In order to retain the particles in sol form it is clearly preferable to keep Q small and to reduce S as far as possible; this may be done in the case of barium sulfate by carrying out the reaction in alco-

holic solution, in which the solubility of the sulfate is much less than in water. A stable sol of barium sulfate, with particles of about 0.1μ diameter, can be obtained by mixing 0.02 N solutions of barium thiocyanate and cobalt sulfate in 50 per cent alcohol. Owing to the sparing solubility of the barium sulfate the particles do not grow at an appreciable rate and the sol is stable.

Sols of substances which are very sparingly soluble in water, e.g., metals, sulfur and certain metallic sulfides, i.e., with S very low, can be readily obtained by starting with a relatively concentrated solution of a suitable compound. Various chemical reactions may be used, provided excess of electrolyte can be avoided, as shown by the following examples.

(i) *Reduction*.—This method is particularly valuable for the preparation of metal sols and is of historical interest as the earliest known gold sols were obtained in this way. The reducing agents which have been employed are hydrogen, phosphine, carbon monoxide, phosphorus, formaldehyde, hydrazine, phenyl hydrazine and hydrogen peroxide; these are all nonelectrolytes and do not produce electrolytes which would cause precipitation to occur. A silver sol may be prepared by passing a current of pure hydrogen through a saturated solution of silver oxide in water at 50° to 60° C., whereas gold sols are readily obtained by reduction of chloroauric acid solution, containing a little potassium carbonate, by formaldehyde, white phosphorus dissolved in ether, tannic acid, or hydrazine. The latter reagent has also been used for the preparation of sols of silver, platinum, palladium, iridium, tellurium and selenium, a small quantity of gum arabic being used as protective colloid (A. Gutbier, 1902–05). Exceptionally stable concentrated sols of a number of metals were obtained by C. Paal (1902–06) using sodium protalbate and lysalbate as both reducing and protective agents (cf. p. 1250). In the method of Carey Lea (1889), silver nitrate is reduced by an alkaline solution of ferrous sulfate and sodium citrate to form a deep red silver sol in which the ferrous-citrate complex exerts a protective effect.

(ii) *Oxidation*.—A good illustration is the formation of sulfur sols by oxidation of aqueous hydrogen sulfide solution with air or sulfur dioxide; the familiar Wackenroder's solution contains a large proportion of colloidal sulfur. Selenium sols may be similarly prepared by oxidation of solutions of hydrogen selenide.

(iii) *Double decomposition*.—The preparation of barium sulfate sol has been already mentioned, and silver halide sols can also be obtained by double decomposition, provided a slight excess of either silver or halide ions is present to act as stabilizer (p. 1241). The passage of hydrogen sulfide through dilute solutions of arsenious oxide or of mercuric cyanide yields sols of arsenious or mercuric sulfide, respectively; in each case the other product is a weak electrolyte, i.e., water or hydrogen cyanide, and so there is little tendency for precipitation.

(iv) *Hydrolysis*.—This method is a form of double decomposition with water as one of the reactants; it is used to obtain sols of oxides, or hydroxides, of weakly electropositive metals, e.g., iron and aluminum, and of some non-metals. A sol of ferric oxide is prepared by adding a small quantity of ferric chloride to boiling water; hydrolysis occurs and ferric hydroxide and hydrochloric acid are produced. The former remains in colloidal solution and the latter is removed by dialysis. Sols of the hydroxides of iron, chromium, aluminum, tin, bismuth, cerium, thorium and zirconium can be obtained by dialyzing solutions of the corresponding nitrates. Silicic acid sols have been prepared by hydrolysis of silicon sulfide, and of methyl and ethyl orthosilicates.

(v) *Exchange of Solvent*.—Although no chemical reaction is involved, the application of (70) shows that when a relatively concentrated solution, i.e., Q is high, of a substance is added to a large volume of liquid in which it is very sparingly soluble, i.e., S is low, the velocity of condensation should be considerable, and particles of colloidal size should result under suitable conditions. Sols of gum mastic, sulfur or phosphorus can be made by pouring alcoholic solutions into a large excess of water. A modification of the same type of method is concerned in the preparation of colloidal silver iodide by pouring into water a solution of the iodide in concentrated potassium iodide solution, in which the silver is probably present as a complex anion.

Dispersion Methods.—Starting with the material in the massive form, it is possible by suitable devices to disintegrate it into particles of colloidal dimensions which remain for some time in the dispersed state.

(i) *Mechanical Disintegration*.—Many substances can be reduced to colloidal size in a "colloid mill," consisting generally of a series of metallic discs, each rotating at very high speed, in a direction opposite to that of its neighbors. The dispersion medium together with the substance to be dispersed, and a protective agent, is passed through the mill and after a time a colloidal solution results. The protective material stabilizes the sol and prevents the particles from coagulating.

(ii) *Peptization by Ions*.—Freshly precipitated substances can often be peptized, i.e., re-dispersed into the primary particles which coagulate to form the precipitate, by suitable reagents, generally one with an ion in common with the material to be dispersed (cf. p. 1241). Freshly precipitated ferric hydroxide, from ferric chloride and ammonium carbonate, can be peptized by ferric chloride; and many sulfide precipitates may be dispersed by passing hydrogen sulfide through the water in which they are suspended. The so-called zincate solution obtained by the action of sodium hydroxide on precipitated zinc hydroxide is partly a colloidal solution of the latter with the hydroxyl ions as the peptizing agent.

(iii) *Washing Methods*.—It is a common experience in analytical chemistry that a precipitate tends to run through the filter paper while being washed free from electrolytes. The latter have caused the primary colloidal particles to form a visible precipitate, and their removal may result in a return to the colloidal state.

(iv) *Electrical Disintegration*.—By passing a direct current arc between wires of platinum, gold or silver immersed in water G. Bredig (1898) obtained sols of these metals. In view of their relative instability electrolytes should be absent, although traces of alkali hydroxide are advantageous (p. 1242). The electrical method was considerably improved and extended by T. Svedberg (1905-09) who succeeded in obtaining organosols, as well as hydrosols, of metals and of nonmetallic substances having a moderate electrical conductivity; he employed alternating current, since this resulted in less decomposition of the organic dispersion medium. The electrodes were generally of iron or aluminum, which do not disintegrate easily, and the material to be dispersed was suspended in the form of granules or foil in the medium through which the arc was passed. Svedberg was of the opinion that the heat of the arc or spark first caused the substance to vaporize and that the colloidal particles were subsequently formed by condensation.

EMULSIONS

The Properties of Oil in Water and Water in Oil Emulsions.—An emulsion * is a disperse system in which disperse phase and dispersion medium are normally liquids; familiar examples of such systems are milk and mayonnaise, consisting of particles of liquid fat dispersed in water, and cod-liver oil emulsion, in which the distribution of phases is reversed. The particles, or globules, of dispersed liquid are from 0.1 to 1 μ , or more, in diameter and so are larger than those found in sols; they are generally visible in the microscope and sometimes to the unaided eye. Emulsions are of considerable technical importance, but they present a very complex theoretical problem that has not yet been completely solved. If a hydrocarbon oil is shaken with water, or if a liquid which is immiscible or partially miscible with water, e.g., aniline, is distilled in steam, or if a solution of the liquid in alcohol or acetone is poured into excess of water, a faintly milky system is often obtained which is a weak emulsion of "oil" in water; the word "oil" is used as most emulsions studied consist of an oil, mineral or vegetable, and water. In the emulsions described the oil is the disperse phase and so they are called **oil in water** emulsions, to distinguish them from the **water in oil** type in which particles of water are dispersed in oil as the continuous medium.

The type of emulsion may be identified experimentally in various ways. If a small quantity of a water in oil emulsion is placed on a microscope slide and a drop of oil stirred into it, mixing will occur readily, but if the emulsion is of the oil in water type the drops will not coalesce. Another method is to add a small amount of an oil-soluble dyestuff to the system; if oil is the continuous phase, i.e., water in oil emulsion, it will become deeply colored, but an oil in water emulsion remains colorless, for the dyestuff is insoluble in water and so cannot reach the dispersed oil. An alternative procedure often employed is to measure the electrical conductivity; as may be expected, this is appreciable when water is the dispersion medium, but negligible if oil is the continuous phase. It was suggested by Walther Ostwald (1910) that the type of emulsion formed depended on the ratio of the volumes of the two phases. It is well known that when spheres are packed as closely as possible they occupy only 74.06 per cent of the total volume, and hence it appears that an emulsion containing more than 74 per cent by volume of the disperse phase should not be possible. A larger relative volume may then lead to a reversal of phase. This view involves the assumption that the drops of liquid of the disperse phase always remain spherical and are never deformed.† Nevertheless, the relative volumes of the two phases probably have some influ-

* The so-called photographic emulsions are not emulsions in the strict sense, but are silver halides dispersed in gelatin gel.

† The so-called emulsion prepared by S. U. Pickering (1907) containing 99 parts of kerosene oil dispersed in one part of soap solution is often quoted as evidence in favor of the deformability of liquid drops and as an argument against the phase-volume theory. It has now been shown (A. S. C. Lawrence, 1937) that the system is not an emulsion but a paste of oil and soap.

ence on the nature of the emulsion formed, particularly in the absence of an emulsifying agent. For instance, if a mixture of oleic acid and water is passed through a colloid mill then an oil in water emulsion results if the mixture contains from 5 to 40 per cent of the oleic acid, but if the proportion is 40 to 70 per cent a water in oil emulsion is obtained (A. J. Stamm and E. O. Kraemer, 1926).

Water in oil emulsions can usually exist only in the presence of a third substance, and the oil in water emulsions obtained by merely shaking the two components together contain only very little disperse phase and are not very stable. In order that an emulsion consisting of a large number of small drops of liquid, and hence possessing a large area of oil-water interface, should be stable, the interfacial tension must obviously be small; otherwise the globules will tend to coalesce and bring about an appreciable reduction in the free energy of the system. The interfacial tension between water and a liquid insoluble in it is generally relatively large (p. 485), and so in the absence of any added material, e.g., soap, it is apparent that only a "dilute" emulsion of oil in water can be obtained. Such systems have properties similar to lyophobic sols. The globules show Brownian movement, provided they are not too large, they carry a negative electrical charge, and can be coagulated by electrolytes. The electrophoretic mobility is of the same order as for colloidal particles (p. 1239), i.e., about 4×10^{-4} cm. per sec. under a fall of potential of 1 volt per cm., and application of (49) shows that the electrokinetic potential, assuming the presence of a double layer, is about 0.05 volt. The addition of electrolytes affects the zeta-potential and the electrophoretic velocity in the same manner as for a lyophobic sol with negatively charged particles; the positive ion is the effective one, and the influence increases markedly with increasing valence. It is probable, although not certain, that the stability of these emulsions is dependent to some extent on the adsorption of ions.

Emulsifiers.—The presence of substances known as **emulsifying agents** permits of the preparation of stable emulsions containing a larger proportion of disperse phase. Such systems have properties which resemble somewhat those of lyophilic sols, e.g., high viscosity, relatively high concentrations, and stability to electrolytes. An excess of electrolyte may salt out the emulsifier and so cause instability. Emulsifying agents fall roughly into three categories. These are first, the long-chain compounds with polar groups, such as the soaps and long-chain sulfonic acids and sulfates, all of which produce a very considerable decrease in the oil-water interfacial tension. It may be noted here that the detergent, i.e., cleansing, action of soap* is generally ascribed to its ability to emulsify grease. When olive oil and water are shaken together very little emulsifi-

* The term "soap," as used here and subsequently, may be taken as including the many detergents which have been prepared and have found important application in recent years. These substances are all characterized by the presence of a polar group and a long hydrocarbon chain, and many of them are colloidal electrolytes.

cation occurs, but the addition of a small amount of sodium hydroxide results in the formation of a stable emulsion; the sodium soap formed by hydrolysis, or by reaction with traces of long-chain acid, acts as the emulsifier. It appears that there is a certain optimum concentration of soap; amounts less or more than this do not cause effective stabilization. Secondly, there are substances of a lyophilic nature, such as proteins, e.g., casein in milk, and gums, and thirdly, various insoluble powders, e.g., basic sulfates of iron, copper or nickel, finely divided lead sulfate and ferric oxide, and lamp black, which stabilize a number of emulsions. The soaps of the alkali metals favor the formation of oil in water emulsions, but those of the alkaline earths, and of zinc, iron and aluminum give water in oil systems. Similarly, the basic sulfates stabilize oil in water emulsions, whereas the opposite type is formed when finely divided carbon is the emulsifying agent. There are a few cases where a soluble substance is able to bring about emulsification, e.g., iodine in the ether-water system.

Reversal of Phase.—The interesting observation has been made that an oil in water emulsion, e.g., hydrocarbon or olive oil in water, with a sodium or potassium soap as emulsifier, may be converted into a water in oil system by the addition of salts of bi- and ter-valent cations, e.g., aluminum, iron, chromium, nickel, lead, magnesium, barium, strontium and calcium (G. H. A. Clowes, 1913; S. S. Bhatnagar, 1920–21). The effect is then the same as if a soap of one of these metals had been employed. The action of the ions is not due merely to their valence, for the amounts necessary show that a chemical reaction is involved. By adjusting the amount of calcium salt added it is possible to obtain an unstable system which is neither a water in oil nor an oil in water emulsion; the addition of a slight excess of sodium ions, however, stabilizes the former, whereas calcium ions stabilize the latter (G. H. A. Clowes, 1916). It has been claimed that this is an illustration of the “antagonistic action” of ions which is encountered in certain biological phenomena, but it is doubtful whether the cases are quite analogous. It must be emphasized that the reversal of phase resulting from the addition of multivalent ions is applicable when a soap is the emulsifying agent, for in other cases different behavior has been observed. A hydrocarbon oil in water system, emulsified by casein, is stabilized by the addition of aluminum sulfate or thorium nitrate; barium chloride has little effect, but the hydroxide is a good stabilizer (W. Seifriz, 1925).

Breaking of Emulsions.—Emulsions may be broken, that is, converted into two macroscopic liquid layers, in several ways.* Chemical destruction of the emulsifying agent, e.g., the addition of acid or a salt of a bivalent cation to an emulsion stabilized by a sodium soap, is effective, as also is the addition of a substance which tends to bring about a reversal of phase. It has been mentioned that high valence ions of opposite sign to the emulsion particles reduce

* The process is sometimes referred to as **demulsification**.

the zeta-potential; they will eventually cause the emulsion to break. Physical and mechanical means are often employed for the same purpose; mention may be made of heating, freezing, centrifuging or sudden jarring. The conversion of cream to butter by churning is an example of the breaking of an emulsion of fat in water.

Theory of Emulsification.—There is little doubt that different causes are operative in the stabilization of different emulsions, although there may well be one fundamental underlying principle applicable to all cases. When the emulsifier is an alkali metal soap its effect is probably due to adsorption at the oil-water interface and consequent reduction of the interfacial tension (G. Quincke, 1888; F. G. Donnan, 1899); as already seen, a low interfacial tension would favor the formation of a stable emulsion with a large oil-water surface. F. G. Donnan and H. E. Potts (1910) have shown that the tension at the interface between a mineral oil and water is lowered by the sodium salts of fatty acids, the lowering for a given concentration being greater the longer the fatty acid chain. The emulsifying effect of the sodium salt also increases with the chain length. With the salts of lower acids the amounts required to reduce the interfacial tension sufficiently to produce emulsification are so large that precipitation actually occurs, and the optimum concentration for emulsification with the higher salts, i.e., the soaps, is ascribed to the opposing influence of precipitation by the ions and stabilization by the reduction of interfacial tension by the soap film.

Type of Emulsion: Orientation Theory.—A more detailed picture of the conditions at the interface has been given by W. D. Harkins, *et al.* (1917). A soap or similar molecule consists of a long hydrocarbon chain and a polar end-group; at an oil-water interface the latter will certainly be oriented throughout towards the water and the hydrocarbon chain will tend to point towards the oil. If the cross sections of the two parts of the stabilizing soap molecule are such that the polar end is larger than the hydrocarbon chain, then a stable emulsion will be favored in which the area of the water side of the soap film is greater than that of the oil side; that is, the water will be the external, or continuous, phase and the oil the internal, or disperse, phase. An oil in water emulsion will thus be stabilized, as is the case with alkali salts of fatty acids. If, on the other hand, the hydrocarbon portion of the soap has a larger cross section than the polar group, the reverse type of emulsion, that is, water in oil, will be stabilized. In the soaps of univalent metals there is only one hydrocarbon chain to each metal ion, but with the soaps of multivalent metals there will be two or more chains for each cation; it is reasonable to suggest, therefore, that the hydrocarbon chains have a larger cross section than the multivalent cation and that this accounts for the stabilization of the water in oil emulsions. This view, sometimes called the "oriented wedge" theory of emulsification, has been tested by J. H. Hildebrand, *et al.* (1923) with a number of oleates and stearates, and the results apparently

support the theory; the larger the univalent cation the more stable is the oil in water emulsion. Although the theory is attractive it meets with difficulties: one is that it is necessary to suppose that the soap film consists of undissociated molecules, the cations being unhydrated, and another is that the curvature of the interface is excessively small in comparison with molecular dimensions. The interface, as far as the molecules of soap are concerned, is virtually flat and it is unlikely that a small change in the cross-sectional area of the soap would result in a reversal of curvature.

Emulsification and Wetting.—A more general theory has been proposed which is capable of development. If the tension between the water and the soap film is smaller than that between the soap film and oil, the former surface will tend to be larger than the latter and an oil in water emulsion will result; on the other hand, if the interfacial tension between the soap film and oil is the smaller, a water in oil emulsion will be obtained.* Further if the interfacial tension between the water and the emulsifier film is less than that between the film and oil, the material of the film will tend to be drawn into the water (cf. p. 484), so that a substance which is preferentially wetted, or, looked at in a somewhat different way, is more easily dispersed, by water will favor oil in water emulsions, whereas an emulsifier capable of being wetted or dispersed by oil will stabilize water in oil systems. It is certainly a fact that calcium and magnesium oleates are dispersed by oils to form colloidal solutions, and these substances favor water in oil emulsions; on the other hand, the alkali metal oleates are dispersed in water and so stabilize the reverse type. Gum dammar is dispersed in oil and it favors water in oil emulsions, but gum arabic is dispersed in water and is frequently employed as an oil in water emulsifier. It appears necessary that the emulsifying agent should be dispersed, and not in true solution, if the generalization made above is to apply. For example, oleic acid is soluble in benzene and amyl alcohol, but not in water, yet it emulsifies the organic liquid as disperse phase in water. Similarly, iodine emulsifies ether in water, although it is more soluble in the former; in these instances the emulsifying agent is in true solution. The views concerning the stabilizing of emulsions, based on surface tension effects, thus appear to be generally applicable, but the theory does not take into account certain factors, such as the influence of the relative amounts of the two liquid phases, which are undoubtedly of importance.

Although it is generally necessary to add a multivalent ion to bring about phase reversal, it may be noted that when the emulsifier is not a very soluble soap, e.g., sodium oleate or stearate, inversion can be brought about by addition of sodium chloride. The latter presumably reduces the

* An aqueous solution of sodium oleate, which stabilizes oil in water emulsions, spreads on olive oil, indicating a low interfacial tension; on the other hand, calcium oleate in aqueous solution does not spread on olive oil, and this favors water in oil systems.

solubility of the soap in the water to such an extent that it becomes relatively more soluble in the oil, and hence a water in oil system is stabilized. Oil in water emulsions containing the soluble potassium soaps do not suffer phase inversion when potassium chloride is added. It is not impossible that in the former instance, that is, with the sodium soap, the emulsifier is actually salted out and a solid film is formed at the interface.

Solid Emulsifiers.—If a given liquid wets a solid more readily than another, that is to say, the tension at the interface between the former liquid is less than that for the latter, then the solid will tend to be drawn into the former. Consequently, when a powder is used to bring about emulsification most of the surface area of the particles will be in the liquid which wets them more readily; this becomes the external, i.e., disperse, phase. Carbon black is more easily wetted by oil than by water; hence it favors a water in oil type of emulsion. On the other hand, silica permits the emulsification of oil in water, since it is more readily wetted by water; it has been stated that if an appropriate mixture of silica and carbon black is used no emulsion is formed because of their opposing action (T. R. Briggs, 1921).

The theory that the stability of an emulsion depends on relative interfacial tensions and, as an approximate corollary, on the dispersibilities of the emulsifier in the two phases, is applicable in many instances, but there are probably several other factors which determine the formation of emulsions in the first place and their subsequent stability in the second place. A complete theoretical interpretation is rendered difficult by the experimental uncertainties: for example, it is not easy to define a "stable" emulsion, and further it is not always certain that any particular emulsion is in a state of equilibrium. Added to these difficulties there is the perplexing fact that with exactly the same components one mode of treatment will yield a relatively stable emulsion, whereas another will not. It is probable, nevertheless, that the theoretical principles developed above provide at least a partial guide to the solution of the problem of the stability of emulsions. Such matters as the relative amounts of the two liquids and especially the electrical charge carried by the dispersed particles, concerning which little is known, are also undoubtedly of importance.³⁰

REFERENCES

- Adam, N. K. "The Physics and Chemistry of Surfaces."
Alexander, J. "Colloid Chemistry."
Bancroft, W. D. "Applied Colloid Chemistry."
Bogue, R. H. "Theory of Colloidal Behavior."
Brunauer, S. "The Adsorption of Gases and Vapors."
Burton, E. F. "The Physical Properties of Colloidal Solutions."
Clayton, W. "Theory of Emulsions and Emulsification."
Freundlich, H. "Kapillarchemie."
Gregg, S. J. "The Adsorption of Gases by Solids."
Hartman, R. J. "Colloid Chemistry."
Hauser, E. A. "Colloidal Phenomena."

- Hedges, E. S. "Colloids."
Kruyt, H. R. "Colloids" (Trans. by H. S. van Klooster).
McBain, J. W. "The Sorption of Gases."
Svedberg, T. "Colloid Chemistry."
Taylor, H. S. "Treatise on Physical Chemistry," 2nd. ed., Chap. XX (E. O. Kraemer).
Thomas, A. W. "Colloid Chemistry."
Weiser, H. B. "Colloid Chemistry."
1. Kayser, *Ann. Phys.*, **14**, 451 (1881); McBain, *et al.*, *Phil. Mag.*, **18**, 916 (1909); *Z. phys. Chem.*, **68**, 471 (1909); *J. Am. Chem. Soc.*, **48**, 690 (1926); **52**, 2217 (1930); *Proc. Roy. Soc.*, **125A**, 579 (1929).
 2. Langmuir, *J. Am. Chem. Soc.*, **38**, 2221 (1916); **40**, 1361 (1918); Laidler, Glasstone and Eyring, *J. Chem. Phys.*, **8**, 659 (1940).
 3. Taylor, *et al.*, *J. Am. Chem. Soc.*, **53**, 578, 2168, 3604, etc. (1931); *Chem. Rev.*, **9**, 1 (1931); Eucken and Hunsmann, *Z. phys. Chem.*, **B44**, 163 (1939); Brunauer, *et al.*, *J. Am. Chem. Soc.*, **60**, 309 (1938); **62**, 1723 (1940); *op. cit.*; Kistler, *et al.*, *ibid.*, **65**, 1909 (1943); Pickett, *ibid.*, **67**, 1958 (1945).
 4. Marshall and Bramston-Cook, *J. Am. Chem. Soc.*, **51**, 2019 (1929); Ganguli, *J. Phys. Chem.*, **34**, 665 (1930); Garner, *et al.*, *J. Chem. Soc.*, 837 (1931); 1436 (1935); 857, 893 (1939); Beebe, *et al.*, *J. Am. Chem. Soc.*, **56**, 642 (1934); **57**, 2527 (1935); **58**, 2196 (1936); **59**, 1627 (1937); **60**, 2912 (1938); **62**, 2134 (1940); Lamb, *et al.*, *ibid.*, **57**, 2154 (1935); Roberts, *Proc. Roy. Soc.*, **154A**, 453 (1935).
 5. Lewis, *Phil. Mag.*, **15**, 499 (1908); **17**, 466 (1909); Donnan and Barker, *Proc. Roy. Soc.*, **85A**, 557 (1911); McBain, *et al.*, *J. Am. Chem. Soc.*, **49**, 2230 (1927); **51**, 3534 (1929); **62**, 1319 (1940); *J. Phys. Chem.*, **36**, 300 (1932); *Proc. Roy. Soc.*, **154A**, 608 (1936); **174A**, 286 (1940); Long and Nutting, *J. Am. Chem. Soc.*, **63**, 625 (1941); Alexander, *Trans. Farad. Soc.*, **38**, 248 (1942).
 6. Langmuir, *J. Am. Chem. Soc.*, **39**, 1848 (1917); Adam and Jessop, *Proc. Roy. Soc.*, **110A**, 423 (1926).
 7. For summary and references, see Adam, *op. cit.*; a shorter account is given in "Recent Advances in Physical Chemistry," by S. Glasstone; see also, Langmuir, *Proc. Roy. Soc.*, **170A**, 1 (1939); Harkins, *Chem. Rev.*, **29**, 385 (1941).
 8. Schofield and Rideal, *Proc. Roy. Soc.*, **109A**, 57 (1925); **110A**, 167 (1926).
 9. Blodgett, *J. Am. Chem. Soc.*, **57**, 1007 (1935); *J. Phys. Chem.*, **41**, 975 (1937); Langmuir, *ref.* (7).
 10. Paneth, *et al.*, *Phys. Z.*, **15**, 924 (1914); *Z. phys. Chem.*, **101**, 445, 480 (1922); *Be.*, **57**, 1215, 1221 (1924).
 11. Zechmeister and Cholnoky, "Principles and Practice of Chromatography"; Strain, "Chromatographic Adsorption Analysis"; Wilson, *J. Am. Chem. Soc.*, **62**, 1583 (1940); DeVault, *ibid.*, **65**, 532 (1943); Cassidy, *J. Chem. Ed.*, **16**, 88 (1939); Eliel, *ibid.*, **21**, 583 (1944); Weiss, *J. Chem. Soc.*, 297 (1943).
 12. Stern, *Z. Elek.*, **30**, 508 (1924); Verwey, *Chem. Rev.*, **16**, 363 (1935); Butler, "Electrocapillarity," Chap. IV; see also, General Discussion, *Trans. Farad. Soc.*, **36**, (1940).
 13. For reviews, see Abramson, "Electrokinetic Phenomena"; Butler, *ref.* (12), Chap. V; Moyer, *Trans. Elec. Soc.*, **73**, 488 (1938); Horwitz, *J. Chem. Ed.*, **16**, 519 (1939); see also, General Discussion, *ref.* (12); Glasstone, "Introduction to Electrochemistry," Chap. XVI.
 14. Lippmann, *Ann. Phys.*, **149**, 546 (1873); *Ann. Chim. Phys.*, **5**, 494 (1875); Frumkin, *7th Colloid Symposium Monograph*, 89 (1930); for review, see Adam, *op. cit.*
 15. Picton and Linder, *J. Chem. Soc.*, **61**, 148 (1892); Burton, *Phil. Mag.*, **11**, 425; **12**, 472 (1906); **17**, 583 (1909); Ellis, *Z. phys. Chem.*, **89**, 145 (1914); Pennycook, *J. Chem. Soc.*, 1447 (1930).
 16. Lottermoser, *J. pr. Chem.*, **72**, 39 (1905); Powis, *J. Chem. Soc.*, **107**, 818 (1915); Young and Goddard, *J. Phys. Chem.*, **21**, 1 (1917); Zsigmondy, "The Chemistry

- of Colloids" (Trans. by E. B. Spear); Pennycuik, *J. Chem. Soc.*, 551, 2018 (1928); *J. Am. Chem. Soc.*, 52, 4621 (1930).
17. Schulze, *J. pr. Chem.*, 25, 431 (1882); 27, 320 (1883); Linder and Picton, *J. Chem. Soc.*, 63, 87 (1895); 83, 1906 (1905); Hardy, *Proc. Roy. Soc.*, 66A, 110 (1900); Whitney and Ober, *J. Am. Chem. Soc.*, 23, 842 (1901); Buxton and Teague, *Z. phys. Chem.*, 57, 76 (1906); Freundlich, *ibid.*, 73, 385 (1910); Weiser, *et al.*, *J. Phys. Chem.*, 24, 30 (1920); 40, 1071 (1936); 42, 427 (1938); *J. Am. Chem. Soc.*, 62, 1924 (1940); Ostwald, *J. Phys. Chem.*, 42, 981 (1938); Fisher and Sorum, *ibid.*, 44, 62 (1940); Hazel, *ibid.*, 45, 731, 738, 747 (1941).
 18. Bjerrum, *Z. phys. Chem.*, 110, 756 (1924); Pauli and Valkó, *ibid.*, 111, 161 (1926); McBain and McBain, *ibid.*, 161, 279 (1932).
 19. For summaries, see Fajans, "Radioelements, Isotopes, Chemical Forces"; Kolthoff, *Chem. Rev.*, 16, 87 (1935).
 20. Zsigmondy, *Z. anorg. Chem.*, 49, 697 (1901); Gortner, *J. Am. Chem. Soc.*, 42, 595 (1920); Hauser and Cincs, *J. Phys. Chem.*, 46, 705 (1942).
 21. Kruyt, *et al.*, *Kolloid. Beih.* 28, 1, 407 (1928); 29, 396, 413, 432 (1929); de Jong, *ibid.*, p. 454; Dhar and Ghosh, *Kolloid-Z.*, 48, 43 (1929); Langmuir, *J. Chem. Phys.*, 6, 873 (1938); Briggs, *J. Phys. Chem.*, 45, 866, 943 (1941). For reviews of the Tiselius method, see Tiselius, *Trans. Farad. Soc.*, 33, 524 (1937); Longworth and MacInnes, *Chem. Rev.*, 24, 271 (1939); Abramson, *et al.*, *ibid.*, p. 345; Longworth, *ibid.*, 30, 323 (1942).
 22. For summaries, see Svedberg, *Chem. Rev.*, 14, 1 (1935); 20, 81 (1937); *Proc. Roy. Soc.*, 170A, 40 (1939); Svedberg and Pedersen, "The Ultracentrifuge"; see also McBain, *et al.*, *J. Am. Chem. Soc.*, 56, 1021 (1934); 59, 2489 (1937); *J. Phys. Chem.*, 42, 1063 (1938).
 23. Walpole, *Kolloid-Z.*, 13, 241 (1913); Laing and McBain, *J. Chem. Soc.*, 117, 1506 (1920); *Nature*, 125, 125 (1930); Weiser, *Colloid Symposium Monograph*, 38 (1924); "Hydrous Oxides"; Hurd, *Chem. Rev.*, 22, 403 (1938).
 24. Donnan, *et al.*, *Z. Elek.*, 17, 572 (1911); *J. Chem. Soc.*, 99, 1554 (1911); *Chem. Rev.*, 1, 73 (1924); *Z. phys. Chem.*, 162, 346 (1932); Procter, *J. Chem. Soc.*, 105, 313 (1914); Procter and Wilson, *ibid.*, 109, 307 (1916); Loeb, "Proteins and the Theory of Colloidal Behavior."
 25. Bachmann, *Z. anorg. Chem.*, 73, 125 (1911); Bogue, *J. Am. Chem. Soc.*, 44, 1343 (1922); Freundlich, *J. Phys. Chem.*, 41, 901, 1151 (1937); Myers and France, *ibid.*, 44, 1113 (1940); Flory, *ibid.*, 46, 132 (1942); *J. Am. Chem. Soc.*, 63, 3083, 3091, 3096 (1941).
 26. Szegvari and Schalek, *Kolloid-Z.*, 32, 318; 33, 326 (1923); Freundlich, *Z. phys. Chem.*, 131, 278 (1928); ref. (25); Russell and Rideal, *Proc. Roy. Soc.*, 154A, 540 (1936); Goodeve, *Trans. Farad. Soc.*, 35, 342 (1939); Myers and France, ref. (25).
 27. Wilhelm Ostwald, *Z. phys. Chem.*, 23, 365 (1897); Bradford, *Biochem. J.*, 10, 169 (1916); Holmes, *J. Am. Chem. Soc.*, 40, 1187 (1918); Wolfgang Ostwald, *Kolloid-Z.*, 36, 380 (1925); Hedges, *ibid.*, 52, 219 (1930); "Liesegang Rings and other Periodic Structures"; van Hook, *J. Phys. Chem.*, 42, 1191, 1201 (1938); 45, 879 (1941).
 28. McBain, *et al.*, *Ber.*, 43, 321 (1910); *Z. phys. Chem.*, 76, 179 (1911); *Trans. Farad. Soc.*, 9, 99 (1913); *Proc. Roy. Soc.*, 97A, 44 (1920); *J. Am. Chem. Soc.*, 42, 426 (1920); 50, 1636 (1928); 57, 1905, 1909, 1913, 1916 (1935); 65, 2072 (1943); *J. Chem. Soc.*, 105, 957 (1914); 113, 825 (1918); 121, 2325 (1922); Reyckler, *Kolloid-Z.*, 12, 283 (1913); Hartley, *et al.*, *Z. phys. Chem.*, 170, 321 (1934); *Trans. Farad. Soc.*, 31, 31, 120, 183 (1935); 32, 795 (1936); Lottermoser and Püschel, *Kolloid-Z.*, 63, 175 (1933); Hoerr and Ralston, *J. Am. Chem. Soc.*, 65, 976 (1943); McBain (M. L.), *ibid.*, 61, 3210 (1939); *J. Phys. Chem.*, 47, 196 (1943).
 29. Giles and Salmon, *J. Chem. Soc.*, 123, 1597 (1923); Highfield, *Trans. Farad. Soc.*, 22, 57 (1926); McBain, *J. Phys. Chem.*, 30, 239 (1926); Soyenkoff, *ibid.*, 35, 2993 (1931); Vernon and Nelson, *ibid.*, 44, 12, 21 (1940).

30. Donnan, *Z. phys. Chem.*, **31**, 42 (1899); *Kolloid-Z.*, **7**, 208 (1910); Bancroft, *J. Phys. Chem.*, **17**, 501 (1913); **19**, 275 (1915); Newman, *ibid.* **18**, 34 (1914); Powis, *Z. phys. Chem.*, **89**, 91, 179 (1914); Ellis, *ibid.*, p. 145; Briggs, *J. Phys. Chem.*, **19**, 478 (1915); *Ind. Eng. Chem.*, **13**, 1008 (1921); Clowes, *J. Phys. Chem.* **20**, 407 (1916); Bhatnagar, *J. Chem. Soc.*, 117, 542 (1920); 119, 61, 1760 (1921); Harkins, *et al.*, *J. Am. Chem. Soc.*, **39**, 595 (1917); **51**, 1674 (1929); Hildebrand, *et al.*, *ibid.*, **45**, 2780 (1923); *J. Phys. Chem.*, **45**, 1303 (1941); Seifriz, *ibid.*, **29**, 739 (1925); Stamm and Kraemer, *ibid.*, **30**, 992 (1926); Tartar, *ibid.*, **33**, 435 (1929); **34**, 373 (1930); King, *Trans. Farad. Soc.*, **37**, 168 (1941); Pink, *ibid.*, p. 180; Clayton, *op. cit.*

INDEX

- Absolute**, reaction rate theory, 1098–1104, 1111
 and activity coefficients, 1111
 and catalysis, 1146
 comparison with collision theory, 1101
 and heterogeneous reactions, 1125, 1145
 and termolecular reactions, 1106
 and unimolecular reactions, 1107
 and viscosity, 501
 temperature scale, 192, 222
 zero, 192, 221, 222
- Absorption and adsorption**, 1194
 bands, 582
 and color, 582–86
 and dispersion of light, 533–34
 and optical activity, 601
 coefficient of gas solubility, 693
 of light, 582
 spectra of liquids, 580–88
 applications, 586–88
 solvent influence, 586
- Acceptor**, electron, 100
- Acetaldehyde**, decomposition, photochemical, 1178
 thermal, 1057, 1084
- Acetic acid**, as solvent, 976
- Acetone**, decomposition, photochemical, 1177
 thermal, 1050
- Acid**, definition, 974–75
 dissociation constant, 977, 980–82
 equilibrium, 975
 neutralization, 983, 1002–1006
 salt, 812
 strength of, 976, 979
 and catalytic effect, 1136
- Acid-base catalysis**, 1087, 1132–40
 general, 1133
 mechanism, 1137
 salt effects, 1138
 specific, 1133
 equilibria, 975
 indicators, *see* Indicators
- Acidic solvent**, 976
- Actinium**, discovery, 118
 series, 134, 136
- Actinometer**, 1045, 1162
- Action constant**, 32
- Activated adsorption**, 1201, 1203
 charcoal, 1194
 complex, 1091
 molecules, 1089
 state, 1091
- Activation**, energy of, 1089, 1090
 calculation, 1092
 and catalysis, 1131, 1146
 determination, 1089
 of heterogeneous reactions, 1125, 1145–46
 and zero-point energy, 1104
 entropy of, 1103, 1110
 free energy of, 1103, 1104
 and catalysis, 1129
 heat of, 1103
 in unimolecular reactions, 1107
- Active centers**, 1149
 deposit, 127
 mass, 816–17, 845, 846
 nitrogen, 1064, 1189
- Activity coefficient**, 684, 960
 of electrolytes, 954–71
 mean, Debye-Hückel equation, 959
 from E.M.F., 964–65
 from freezing points, 689, 965–66
 in gases, 301, 304
 in rate equations, 1114–16, 1139
 rational, 960
 from solubility, 961
 in solution, 684–85
 stoichiometric, 961
 and temperature, 686
 from vapor pressure, 687
 and equilibrium constant, 825
 of gas, 301
 and reaction rate, 826, 1111
 solubility product, 962
 in solution, 683–90
 and standard state, 684
- Adamantine crystals**, 388

- Additive properties, 524
Adhesion, work of, 484-85
Adiabatic calorimeter, 208-209
 change, 198-99, 219-20
 demagnetization, cooling by, 442-43
 expansion, cooling by, 441
Adsorbent, area of, 1203, 1218
Adsorption, 1194-1210
 activated, 1201, 1203
 and catalysis, 1146, 1149
 chromatographic, 1219
 and critical temperature, 1197
 of gases, 1194-1204
 activated, 1201, 1203
 chemisorption, 1201, 1203
 isosteres, 1195
 pressure effect, 1195
 rate determining, 1124
 temperature effect, 1195
 theories of, 1198, 1202, 1203
 types of, 1200
 van der Waals, 1201, 1202
 Gibbs equation, 1206-08
 heat of, 1195, 1201, 1204
 indicators, 1248
 isotherm, classical, 1196, 1200
 Langmuir, 1198
 of multimolecular layers, 1202
 negative, 1205, 1216
 on precipitates, 1218
 and precipitation of colloids, 1245
 from solution, 1215-19
 and stability of colloids, 1241
 at surface of solution, 1205-09
 and surface tension, 1205, 1215
 of unimolecular layers, 1198
Aerosols, 1234
Affinity, 211, 815
 coefficient, 817
 and free energy, 852
 history, 815
Alcohols, association, 115, 509
 hydrogen bonds in, 115
 structure of liquid, 509-10
Aldehydes, photolysis, 1177, 1178
Alkali halides, crystal structure, 386
 lattice energy, 409
 spacing in crystal, 506
 metals, spectra, 59-60
Allotropy, 350
Alloys, structure of, 379-81
 study of, 745-71
Alpha-particle, 119
 charge of, 120
 charge to mass ratio, 119
 disintegration by, 161-62
 and helium, 121
 and induced radioactivity, 168
 ionizing power, 119, 123
 mass of, 120
 in nucleus, 30-31
 passage through matter, 123
 production in artificial disintegration, 162-65
 in radioactivity, 119
 range, 122
 scattering, 23-24, 124
 scintillations, 121
 stability, 31
 stopping power, 123
 structure, 30
 velocity, 122
Alternation law, in spectra, 66, 68
Alumina sol, 1243
Alums, structure of, 398
Amides, association, 115
Ammine, vapor pressure, 781, 784
 dissociation, 847
Ammonia equilibrium, 839
 photolysis, 1176, 1181
Ammonium salts, dissociation of, 321, 848-49
Ampere, 885
Amphiprotic solvent, 977, 978
Ampholyte ions, 996
Ampholytes, 995-99
 dipolar ion theory, 996
 dissociation constants, 997
 hydrogen ion concentrations, 997
 isoelectric points, 998
Amphoteric electrolytes, *see* Ampholytes
Andrews, critical state, 428-29
Angular momentum, of electron, 43, 55, 63, 67
Anharmonicity constant, 567
Anions, 885
Anisotropic liquids, 514
 particles, 1264
 substance, 346
Anistropy factor in optical rotation, 606

- induced optical, 603
- term in Kerr effect, 558
- Anode, 884**
 - irreversible phenomena at, 1027
 - oxidation at, 1035, 1036, 1039
 - rays, 2
- Anthracene, photochemical dimerization, 1168, 1184**
- Anticathode, 3**
- Antimony electrode, 994**
- Anti-Stokes lines, in fluorescence, 1155**
 - in Raman spectra, 577
- Antisymmetric functions, 79, 274**
 - vibrations, 570
- Antonoff's rule, 485**
- Apparent molar properties, 239**
- Aprotic solvent, 975, 977, 979**
- Area of cross section of molecules, 399, 508**
 - on surfaces, 1210, 1212, 1218
- Arrhenius, equation, 1088**
 - theory of activation, 1089
 - electrolytic dissociation, 887
- Arsenious sulfide sol, 1242**
 - precipitation of, 1244, 1247
- Artificial disintegration, *see* Disintegration**
 - radioactivity, *see* Radioactivity
- Association, *see* Molecular association**
- Asymmetry effect in electrical conductance, 904**
- Atomolysis, 255**
- Atom, effective charge, 39**
 - electronic arrangement, 80-92
 - ground state, 54
 - nuclear theory of, 22-25
 - polarization, 536, 548, 549
 - structure of, 22-92
- Atomic bomb, 174**
 - disintegration, *see* Disintegration
 - distances, by electron diffraction, 594, 595
 - by spectra, 579
 - by X-rays, 383
 - energy levels, 43
 - frequency in solids, 421
 - heat capacity, of solids, 413-15, 416-21
 - and characteristic temperature, 419-20
 - at constant volume, 417
 - temperature variation of, 414
 - theory of, 416-21
- nucleus, 23**
 - charge of, 24, 25, 28
 - and group displacement law, 137
 - constitution of, 28, 30, 137, 174
 - dimensions of, 24
 - disintegration of, *see* Disintegration
 - scattering of α -particles by, 23
- number, 25**
 - and nuclear charge, 25
 - and X-rays, 25, 70
 - effective, 111
- orbitals, 101**
- parameters in crystals, 360**
- radii, 382-83**
- scattering factor, 358, 592**
- spectra, *see* Spectra**
- Attraction, forces of, in crystals, 407-09**
 - molecular, *see* Molecular attraction
- Autocatalysis, 1061, 1078**
- Autoxidation, 1143**
- Auxochromic group, 584**
- Avogadro's law, 193**
 - deduction of, 252
 - deviations from, 247
- Avogadro's number, 1**
 - ✓ determination, by Brownian movement, 255-58
 - by diffusion of granules, 259-63
 - from radioactivity, 122, 129
 - by X-ray method, 13
 - value of, 13, 263-64
- Azeotropic mixtures, 719**
- Azimuthal quantum number, 55-58**
 - of molecule, 572
- Balmer series, 37-38, 49-50**
 - of deuterium, 147
- Band spectra, 563, *see also* Spectra, molecular**
 - group in, 564
 - head in, 571
 - progression in, 571
 - sequence in, 571
 - system, 564
- Barger's micro-method for molecular weights, 633**

- Base, definition of, 975
dissociation constant of, 978, 981
equilibrium, 977
neutralization, 983, 1002-06
strength, 977, 979
and catalytic effect, 1136
- Basicity, determination of, 901
- Basic salt, 812
solvent, 975, 983
- Bathochromic effect, 583
- Beckmann method for depression of
freezing point, 647
for elevation of boiling point, 637
thermometer, 637
- Beer's law, 588, 1169
- Benzene, resonance energy of, 590
structure of, 112
structure of crystal, 403
of liquid, 508-09
- Bergmann series, 40
- Berthelot equation of state, 295
applications, 313, 328, 868
- Beta-particles, 119, 124-25
charge to mass ratio, 124
and electrons, 124
in induced radioactivity, 168-171
spectrum of, 124
velocity and penetrating power, 124
- Beta (β) phase in alloys, 380-81
- Bimolecular reactions, in gas phase,
1057, 1060
in solution, 1057, 1112
- Binodal curve, 793
- Bi-radicals, magnetism, 619
in phosphorescence, 1159
- Birefringence, *see* Double refraction
- Bjerrum's treatment of electrolytes,
969
- Blagden's law, 642
- Bohr magneton, 65, 616
- Bohr's theory of atomic spectra, 43-
50
limitations and objections, 45, 58,
72
and wave mechanics, 73, 76, 77
theory of atomic structure, 82
- Boiling point, and critical tempera-
ture, 456
elevation of solutions, 633-42
of colloidal electrolytes, 1267
determination, 637-42
molar, 637
and molecular weight, 637
and osmotic pressure, 672
and heat of vaporization, 456-57
of mixtures, 715, 721
of partially miscible liquids, 732
and pressure, 450, 458-59
systems of constant, 719, 721
systems of maximum, 720
systems of minimum, 718
and vapor pressure, 445
- Boltzmann constant, 31, 229, 264
distribution, *see* Maxwell-Boltz-
mann law
- Bomb, atomic, 174
calorimeter, 210
- Bond angles, by dipole moment, 545
by electron diffraction, 594
by wave mechanics, 105
distances, *see* Bond radii
energy, 588
and heat of formation, 588
and resonance, 590
odd-electron, 106-07
radii, 382-85
additivity, 382, 384
by electron diffraction, 593
and resonance, 111, 392, 403, 595
by spectra, 569, 579
strength, 588
see also Linkage, and Valence
- Bonding electrons, 106
- Boric oxide glass, 519
- Born-Haber cycle, 410
- Bose-Einstein statistics, 273
- Boyle point, or temperature, 246
calculation, 291-92, 439
- Boyle's law, 191
deduction of, 251-52
deviations from, 244
- Brackett series, 39, 49-50
- Bragg equation for X-ray diffraction,
356-57, 505
method of crystal analysis, 365-71
- Brønsted-Bjerrum equation, 1115
- Brønsted, definition of acid and base,
974-75
catalysis relationship, 1136
- Brownian movement, 255, 1237
and Avogadro number, 255-63
- Budde effect, 1166

- Buffer action**, 1006
 solutions, 1006-08
 table of, 1007
 "universal," 1008
- Buoyancy microbalance**, 309
- Bury theory of atomic structure**, 82
- Cadmium chloride**, lattice type, 389, 390
 iodide, lattice type, 388-90
- Cailletet and Mathias law**, 431
- Calcium chloride-water system**, 775
- Calomel electrode**, 941, 942, 993
- Calorie**, definition of, 182, 184
- Calorimeters**, 207-11
- Canal rays**, 2
- Capillarity**, 487
- Capillary action**, 487
 rise and surface tension, 486-87
- Carbonate**, dissociation of, 779, 780, 847, 848, 864
 ion, structure of, 112, 392
- Carbon atoms**, chains of, 401
 dioxide, structure of, 112, 551, 578, 590, 595
 monoxide, structure of, 528, 578
- Carboxylic acid**, association of, 115, 319, 321, 400, 510, 680
 structure of liquid, 510
- Carnot cycle**, 219-22, 222-23
 theorem, 218-19
- Catalysis**, 1126-54
 and absolute rate theory, 1145
 acid-base, 1087, 1132-40
 active centers in, 1149
 and absorption, 1147, 1149
 criteria of, 1127-28
 dual theory of, 1132
 and equilibrium, 1127
 and free energy of activation, 1129
 function of surface in, 1147
 by halide ions, 1054, 1140
 heterogeneous, 1144-54
 homogeneous, gaseous, 1129-32
 in solution, 1132-41
 by hydrogen ions, 1052, 1132-39
 by hydroxyl ions, 1054, 1132-40
 intermediate compounds in, 1131, 1150
 negative, 1126, 1142, 1143
 poisoning in, 1151
 promoters in, 1152
 salt effects in, 1138
 simultaneous, 1087, 1132-40
- Catalyst**, 1126
 area of, 1146
 mixed, 1152
- Catalytic coefficients**, 1087, 1133, 1134-37
 and dissociation constant, 1134
 poisons, 1151
- Cataphoresis**, *see* Electrophoresis
- Cathode**, definition, 884
 phenomena at, 1017-27
 rays, 2, 3
- Cathodic deposition of metals**, 1022
 evolution of hydrogen, 1017-22
- Cations**, 885
- Cauchy equation**, 533
- Cell constant**, 892
- Cells**, chemical, 934, 936
 concentration, amalgam, 932
 gas, 933
 without transference, 926
 with transference, 928
 reversible, 921
- Cellulose**, structure of, 404-05
- Cesium chloride**, lattice type, 385-87
- Chain reactions**, 1080-85, 1111
 and free radicals, 1084, 1177, 1179
 in hydrogen-chlorine reaction, 1080, 1174
 in hydrogen-oxygen reaction, 1082
 kinetics of, 1083
 and negative catalysis, 1142
 in photolysis, 1177, 1179
 stationary and nonstationary, 1083
 types of, 1081
- Chapman's equation**, 283
- Characteristic temperature of solid**, 419-22
- Charge on colloidal particles**, 1240
 effective, of atom, 39
- Charles's law**, *see* Gay-Lussac's law
- Chemical cell**, 934, 936
 constant, conventional, 863
 true, 860
 equilibrium, *see* Equilibrium
 kinetics, 1044-1190; *see also* Reactions

- potentials, 238
 - of electrolytes, 962, 965
 - of gases, 301, 306
 - of phases in equilibrium, 477
 - of solutions, 668, 683, 684, 708, 736
 - of surface phase, 1206
- Chemiluminescence, 1097, 1188–90
 - sensitized, 1190
- Chemisorption, 1201, 1203
- Cholesteric phases, 516
- Chromatographic adsorption, 1219
- Chromophoric centers and optical rotation, 602
 - electron, 582
 - and optical activity, 601, 602, 608
- groups, 582
- Circular dichroism, 605
 - polarization, 597
 - and optical rotation, 598
- Clapeyron-Clausius equation, 450–53, 459, 460, 463, 625, 635, 644, 649, 695, 862
- Classical adsorption isotherm, *see* Adsorption
 - distribution law, *see* Maxwell-Boltzmann law
- Clausius gas equation, 295
 - and critical constants, 438
- Clausius-Mosotti equation, *see* Mosotti-Clausius equation
- Cloud-track apparatus, 20
 - photograph, of positron, 20
 - of α -particles, 123
 - of β -particles, 125
- Coacervation, 1254
- Coagels, 1236
- Cohesion pressure, 289
 - and molecular attraction, 299
 - work of, 483
- Colligative properties, 524, 625, 672
- Collision diameter of molecules, 275–77, 282, 284
 - numbers, 276–78, 1096
 - theory of reaction rate, 1095, 1101–02
 - failure of, 1097
 - of unimolecular reactions, 1109
- Collisions, elastic, 50
 - in gases, 276–78
 - inelastic, 51, 1156
 - of second kind, 1156
 - triple, 277
 - with walls, 278
- Colloidal, electrolytes, 1266–70
 - platinum as catalyst, 1153
 - state, 1231–1275
 - systems, classification of, 1234
 - dialysis and filtration, 1232
 - electrical properties, 1238, 1252
 - gelation of, 1236
 - general properties, 1232
 - mutual action of, 1248
 - optical properties of, 1233, 1235
 - in organic media, 1270
 - origin of charge, 1240–42
 - precipitation of, 1233, 1243, 1254
 - mechanism of, 1245
 - preparation of, 1272–76
 - protection of, 1249
 - sensitization of, 1249
 - shape and color, 1238
 - size of particles, 1232, 1237
- Color, effect of groups on, 582–83
 - of organic compounds, 582–86
 - resonance and, 584–85
- Combination principle in spectra, 38
- Complex ions, 972–76
 - electronic structure of, 110–11
 - instability constant of, 973
 - paramagnetism of, 620–22
 - and transference number, 919
 - reactions, 1065
- Component, definition, 474
- Compressibility coefficient, 235, 409
 - and heat capacities, 235, 310, 421
- Compound formation in solution, 677, 713
- Compton effect, 34
- Concentration cells, *see* Cells
 - polarization, 1014, 1026
- Condensation in pores, 1197
- Condensed surface films, 1212
 - system, 744
- Conductance of electrolytes, *see* Electrolytes
 - equivalent, 889
 - and hydrolysis, 989
 - ionic, 894
 - measurement, 889
 - applications of, 899–901
 - minimum, 896, 969

- molar, 889
- ratio, 902
- and solubility, 899
- specific, 889
- water, 891
- Conductometric titration, 899
- Congruent melting point, 755
- Conjugate acid and base, 975
- solutions, 722
- Conjugated double bonds, refraction, 530, 534
 - effect on color, 583
- Consecutive reactions, 1075-85
- Conservation of energy, law of, 184
 - of mass and energy in atomic disintegration, 165-66
- Consolute temperature, 725
 - lower, 726
 - upper, 726
- Constant boiling mixtures, 719, 721
 - heat summation law, 204
- Constitutive properties, 526
- Contact angle, 482, 486
- Continuity of state, 432-34
- Convergence limit, in atomic spectra, 38
 - in molecular spectra, 575
- Cooling curves, 747-49, 762
- Coordinate bond, 100
- Coordinated hydrogen, *see* Hydrogen bond
- Coordination compounds, 110-11
 - structure of, 396-97
 - number in crystal lattice, 375
 - and bond radius, 383
 - and lattice type, 385, 388
- Corresponding states, 440
 - law of, 439-40
 - temperature, boiling point as 456, 524
- Corrosion of metals, 1033, 1035
- Cotton effect, 605
- Cottrell method for boiling point, 640
- Coulomb, 885
- Coulombic energy, 102, 1093
 - in crystals, 407-10
- Coulometer, 886
- Counter-ions, 1247, 1269
- Covalence, 98-100
 - and wave mechanics, 103
 - coordination, 100
- dative, 100
- semipolar, 100
- Co-volume, 289, 500
- Craft's rule, 459
- Cristobalite structure, 388, 390
- Critical complex, *see* Activated complex
 - constants, 429, 432, 437
 - and Berthelot equation, 295
 - and Dieterici equation, 439
 - and limiting density, 314
 - and van der Waals equation, 436
 - opalescence, 726
 - phenomena, 428-29
 - point of binary system, 793
 - potential, 50
 - calculation of, 53
 - measurement of, 50-53
 - pressure, 429-30, 432
 - solution temperature, 725, 797
 - state of gas, 428
 - and equations of state, 438-39
 - indefiniteness of, 434
 - and van der Waals equation, 434-37
 - temperature, 429-30, 432
 - and adsorption, 1197
 - and boiling point, 456
 - and Boyle point, 438-39
 - and surface tension, 492, 494
 - and vapor pressure, 450, 454
 - volume, 429, 431-32, 437
- Crookes dark space, 2
- Cryohydrates, 772
- Cryohydric point, 772
- Cryoscopic constant, 645
- Cryoscopy, 642
- Crystal angles, 340
 - classes, 344
 - coordination number, 375
 - double refraction of, 346
 - elements of, 343
 - form, 341
 - habit, 340
 - lattice, 350-55
 - dimensions, 369
 - energy of, 400-04
 - optic axis of, 346
 - point groups, 336
 - structure, 332-99
 - of elements. 375-77

- experimental methods, 365-67
 - by Fourier synthesis, 362-63
 - of inorganic compounds, 382-99
 - of metals, 377-81
 - of organic compounds, 399-407
 - by Patterson-Harker synthesis, 364
 - of silicates, 394-96
 - by trial analysis, 361
 - X-rays and, 355-407
- symmetry elements of, 341, 354
- systems, 344-45
- zone, 341
- Crystalline state, 340-50
- Crystallization, water of, 397
- Crystallographic axes, 342
 - axial ratios, 343
 - parameters, 342
 - point groups, 344
 - systems, 344
 - and space groups, 355
- Crystals, internal structure, 350-55
 - isomorphous, 346-50; *see also* Isomorphism
 - liquid, 514-17
 - optical properties, 346
- Cube, symmetry elements, 341-42
- Curie point, 611
- Curie's law, 611, 616
- Cybotactic groups, 511
- Cycle or Cyclic process, 217-18
- Cyclotron, 164
- Dalton's law of partial pressures**, 302-03, 305, 306-07
- Daniell cell, 216, 934, 950
- Dative bond, 99
- de Broglie, matter waves, 15, 73, 75
- Debye characteristic temperature, 419
 - equation for dipole moments, 545
 - theory of specific heats, 418-20
 - unit, 547
- Debye-Hückel, complete equation, 966-68
 - limiting law, 960
 - and dissociation constants, 970
 - tests of, 960
 - and weak electrolytes, 969
 - theory of conductance, 903
 - treatment of electrolytes, 956-59
 - extension of, 968
- Decomposition, sensitized, 1157
- voltage, 1014
 - of acids and bases, 1015
 - of aqueous solutions, 1033
 - of salts, 1017
- Degree of dissociation, of electrolyte, 888, 901-02, 907
 - thermal, 319
- Degrees of freedom, of energy, 329
 - of system (phase rule), 475
- Deliquescence, 789
- Demulsification, 1277
- Density of gases, *see* Gas density
- Depolarization factor of light, 558
 - of metal deposition, 1023
- Depolarizer, 1014
 - influence in corrosion, 1034
- Deposition of metals, 1022
 - potential, 1015
- Desensitization of photographic plates, 1187
- Detergents, 1276; *see also* Colloidal electrolytes, and Emulsifying agents
- Deuterium, as indicator, 175, 178
 - discovery of, 147
 - ortho and para states, 97
 - oxide, 160
 - forms of solid, 473
 - properties of, 159
 - rotation in solid, 423
 - separation by chemical methods, 158-59
 - by electrolysis, 156-57
 - factor, 160
 - spectrum of, 147
 - zero-point energy of, 160
- Deuteron, 159
 - disintegration by, 164-65, 166
 - disintegration of, 165
 - and induced radioactivity, 169
 - product of disintegration, 164
- Dew-point and vapor pressure, 632, 1266
- Dialysis, 1232
- Diamagnetic susceptibility, 612
 - additivity in organic compounds, 614
 - of ions, 614
- Diagnetism, 610, 612-15
- Diameter of molecules, *see* Molecular diameter

- Diamond, crystal structure of, 375
heat capacity of, 414
- Diatomic molecule, entropy of, 875
- Dielectric constant and conductance, 896, 897
and dipole moment, 545
measurement of, 545-46
and refractive index, 536
and rotation in solids, 423
- Dieterici equation of state, 295-96
and critical constants, 438-39
- Differential, complete (or exact), 186-87
heat of dilution, 242, 626
heat of solution, 241, 695
oxygenation, 1035
- Diffraction of electrons, 16-19, 591-96
of X-rays, 13, 355-65
- Diffuse double layer, 1221
series of spectra, 40, 41-42, 60, 66
spectra of molecules, 1166
- Diffusion coefficient, 260
current, 1027
Einstein equation of, 260-61
Fick's law of, 1257
flame, 1097
of gases, 254-55, 281
in gels, 1263
Graham's law of, 254
layer, 1026
and molecular diameter, 281-82, 285
potentials, 931
separation of isotopes by, 153-54
in suspensions, 259-62
- Dilution law, 955
- Dipolar ions, 996
- Dipole association, 550
moment, 108-09, 543-56
Debye equation for, 543-45
determination of, 546-49
of groups, 552
and Kerr constant, 558, 559-60
of linkages, 552
and molecular structure, 555
and resonance, 556
results, 551
in solution, 548
solvent effect, 549
vector addition of, 554
term in Kerr constant, 558, 559
- Discharge potential, 1015
- Disintegration of atoms, artificial, 161-67
by α -particles, 161-62
by γ -rays, 165
by deuterons, 164-65
by neutrons, 162-63
by protons, 163-64
conservation of mass and energy in, 165-66
- Disperse phase, 1234
system, 1234
- Dispersion forces in adsorption, 1197
in molecular attraction, 299
of light, 533
medium, 1234
- Dispersivity, molar and specific, 534
- Dissociation constant, of acids and bases, 977-978
and catalytic coefficient, 1136
determination of, 980-82
fundamental, 979
and strength, 980
of ampholytes, 997
of electrolytes, 954
and Debye-Hückel theory, 970
of indicators, 999-1001
of ammonium salts, 321, 848-49
of carbonates, 779, 780, 847, 848, 864
of compounds, 321-22
at melting point, 757
electrolytic, degree of, 888, 901, 907
incomplete, 907
theory of, 887
pressure, 779, 845, 847
thermal, 319
and density, 319-20
of compounds, 321-22
of elements, 320
- Dissymmetry factor in optical activity, 606
- Distillation fractional, 717
in steam, 734
of binary mixtures, 715-21
homogeneous, 717-21
partially miscible, 732
- Distribution law, 697, 735
method for study of association, 737
of complex formation, 739-41

- of hydrolysis, 990
- ratio, 697, 735
- Donnan equilibrium, *see* Membrane equilibrium
- Donor molecule, 100
- Dorn effect, 1220
- Double layer potential, *see* Zeta-potential
 - structure of, 1221, 1227
 - refraction of crystals, 346
 - of liquid crystals, 514, 515
 - streaming, 1264
 - salts, formation, 804–812
- Draper effect, 1173
- Drop-ratio method for pH, 1002
- Drop-weight and surface tension, 689–90
- Dual ions, 996
 - theory of catalysis, 1132
- Dühring's rule, 455
- Duhem-Margules equation, 701, 707, 708
- Dulong and Petit's law, 413–14, 416
- Du Noüy tensiometer, 490
- Dupré equation, 482–83, 484
- Dynamic isomerism, 752
- Dyne, 181
- Ebullioscopic** constant, 636
 - method for molecular weights, 637
- Effective atomic number (E.A.N.), 111
 - charge of atom, 39
- Efficiency of machine, 218–21
- Efforescence, 789
- Effusion, 278–79
 - and vapor density, 279, 318
 - and vapor pressure, 449
- Eigenfunctions, 76; *see also* Wave functions
- Eigenvalues, 76, 77
- Einstein diffusion equation, 260–61
 - extension of quantum theory, 33
 - law of photochemical equivalent, 1159
 - mass-energy relationship, 7
 - theory of specific heats, 417
 - unit of radiation, 1160
- Electrical and chemical energy, 923
 - discharge through gases, 1–3
 - double layer, 1220
 - effect of ions on, 1227
 - units, *see* Units
- Electrocapillary curve, 1229
 - influence of ions, 1229–30
 - influence of nonelectrolytes, 1230
 - phenomena, 1228–30
- Electrochemical passivity, 1028–29
- Electrode, 884
 - irreversible, 1014
 - oxidation-reduction, 949
 - potential, 937
 - determination from entropies, 1013
 - influence of concentration and valence, 947
 - mechanism of, 947
 - sign of, 940
 - significance of, 949
 - single, 937
 - and solution pressure, 947
 - standard, 937, 944–47, 950
 - reference, 941
 - reversible, 922
- Electrodeposition of metals, 1022–26
 - simultaneous, 1025
- Electrodialysis, 1232
- Electrokinetic effects, 1220
 - influence of ions on, 1225
 - quantitative treatment, 1222–25
 - potential, *see* Zeta-potential
- Electrolysis, 1013–40
 - laws of, 885
 - significance of, 886
 - phenomena of, 884
 - separation of isotopes by, 156–57
 - separation of metals by, 1024
- Electrolytes, 884
 - amphoteric, 995
 - application of law of equilibrium, 954
 - colloidal, 1266–70
 - conductance of, 888
 - equivalent, 889, 892
 - effect of concentration, 892
 - effect of temperature and pressure, 895
 - effect of viscosity, 897
 - at infinite dilution, 893
 - at high frequencies and high voltages, 908
 - measurement, 889
 - molar, 889

- solvent influence, 890
- dissociation constant of, 954, 977
- degree of, 888, 901, 907
- incomplete, 907
- theory of, 887
- effect on colloids, 1239-48
- equilibrium in, 954-1000
- interionic attraction theory of, 903, 956
- strong, 892
- thermal properties of, 968, 1009
- and Debye-Hückel theory, 968
- weak, 892
- Electrolytic conduction, 884-921
 - in colloidal electrolytes, 1267-69
 - mechanism of, 887
- conductors, 884
- dissociation, theory of, 887
- oxidation, 1035-40
 - of anions, 1036
 - nonreversible, 1040
 - reversible, 1035
- polarization, 1013
 - anodic, 1027
 - due to concentration changes, 1014, 1026
 - and diffusion, 1026
 - in hydrogen ion discharge, 1017-22; *see also* Overvoltage
 - in metal deposition, 1023
- reduction, reversible, 1035
- nonreversible, 1039
- separation of metals, 1024
- Electromagnetic units, 6-7
- Electrometric titration, *see* Conductometric, *and* Potentiometric
- Electromotive force (E.M.F.), 921-54
 - of chemical cells, 936
 - of concentration cells, 927-30
 - and equilibrium constant, 936, 950
 - measurement of, 924
 - in nonaqueous media, 953
 - sign of, 935
- Electron, 1-2, 4
 - acceptor, 100
 - affinity of halogens, 411
 - angular momentum of, 43, 55, 63, 67
 - antibonding, 106
 - arrangement in elements, 81-92
 - bonding, 106
 - charge of, 8-13
 - by indirect method, 13
 - by oil-drop method, 10-12
 - value of, 13
 - charge to mass ratio, 3-5
 - chromophoric, 582
 - coupling, 67-69
 - density in crystals, 363-64
 - diffraction by crystals, 16-17
 - by gases and vapors, 591-96
 - donor, 100
 - duplet, 99, 103
 - energy levels, 43
 - of sodium, 60
 - extranuclear, 23, 35-92
 - and Faraday, 1, 13
 - groups and subgroups, 81-84
 - lone-pair of, 100
 - magnetic properties of, 65, 611, 617
 - mass of, 6, 13, 48
 - mass and velocity, 7
 - microscope, 18
 - mobility in metals, 379
 - nonbonding, 106
 - octet, 99
 - optical, 43
 - two or more, 66
 - orbits, 43-44
 - and waves, 74
 - elliptic, 55
 - penetrating, 58
 - precession of, 57, 63
 - polarization, 536, 537, 545, 548
 - promoted, 105
 - quantum numbers, 47, 50, 55, 56, 60, 61, 62
 - and Pauli principle, 80
 - radius, 14
 - shells, *see* Electron groups
 - spin of, 61
 - and paramagnetism, 617
 - valence, 43
 - volts (e.v.), 6
 - wave length, 15-16
 - nature of, 14-19
- Electronegativity, 108, 552, 591
- Electronic spectra, 563, 571-76
 - and isotopes, 147, 572
 - of liquids, 580
 - transitions in, 574
 - theory of valence, 97-116

- work function, 33, 949
- Electroosmosis, 1219-20
 - influence of ions, 1225
 - valence rule, 1226
- Electrophoresis, 1238
 - influence of electrolytes, 1239-40
 - Tiselius method, 1252
- Electrophoretic effect in conductance, 904
 - mobility of colloids, 1239
 - of emulsions, 1276
- Electrovalence, 98, 386
 - and homopolar terms, 109
 - transition to covalence, 109, 386, 389, 540
- Electroviscous effect, 1251
- Elements, crystal structure of, 375-77
 - diamagnetism of, 613
 - dissociation of, 320-21
 - heat capacity of, 413-22
 - molecular weights, 318
 - liquid, structure of, 506-07
 - paramagnetism of, 618
- Elliptical polarization, 597
- Ellipticity of polarized light, 605-06
- Emanation, 126-27
- Emulsification, and wetting, 1279
 - theory of, 1278
- Emulsifying agent, 1276, 1277
 - soap as, 1276
 - solid, 1280
- Emulsions, 1234, 1275-80
 - breaking of, 1277
 - interfacial tension theory, 1278
 - oriented wedge theory, 1278
 - phase volume theory, 1275
 - photographic, 1186, 1275
 - reversal of phase in, 1277
 - speed of particles in, 1276
 - types of, 1275, 1280
 - wetting theory of, 1279
- Enantiotropy, 469
 - and monotropy, 471
- Endosmosis, *see* Electroosmosis
- Endothermic reaction, 200
 - rate of, 1092
- Energy, 181
 - activation, *see* Activation
 - chains, 1082
 - changes, 187-88
 - conservation of, 184-85
 - of crystal lattice, 407-13
 - electrical and chemical, 923
 - equipartition principle, 329, 416
 - internal (or intrinsic), 186
 - levels, atomic, 43, 76, 77
 - of sodium, 61
 - of molecules, 563-65
 - and mass relationship, 7
 - of oscillator, 32
 - rotational, 566
 - units, 181
 - vibrational, 567
 - see also* Free energy, and Internal energy
- Enthalpy, *see* Heat content
- Entropy, 222-29
 - at absolute zero, 859, 864, 871
 - of activation, 1103, 1110
 - changes at constant pressure or volume, 226-27
 - at equilibrium, 233
 - in irreversible processes, 227, 233
 - definition, 224
 - determination by third law, 866
 - of fusion, 462-63, 867
 - of gases, 878
 - of ideal gas, 224
 - internal, 872, 874
 - of ions, 1011-12
 - of liquids, 869
 - and magnetization, 442
 - of mixing, 225-26, 865
 - of mixtures, 225, 865
 - and partition functions, 870-72
 - physical significance of, 229
 - and probability, 229, 871
 - and randomness, 229, 442
 - rotational, 875
 - and rotation in solids, 424
 - standard, 869, 878
 - by statistical methods, 870-72
 - and restricted rotation, 877
 - translational, 872, 874
 - of vaporization, 457-58, 867
 - vibrational, 876
 - virtual, 877
- Enzymes, 1128, 1153
- Eötvös equation, 492
- Epsilon (ϵ) phases in alloys, 380-81
- Equation of state, 193, 288-98
 - Beattie and Bridgeman, 298

- Berthelot, 295-96
Clausius, 295
Dieterici, 296-97
ideal gas, 193
Keyes, 297
reduced, 439
thermodynamic, 235-36, 287, 479
van der Waals, 289
virial, 298
- Equilibrium box, 818
and catalyst, 1127
chemical, law of, 817
derivation, 818-21
combination, 841, 851
and concentration cells, 933
constant, 817, 821
activities in, 825
and catalyst, 1127
classical, 826, 955
and E.M.F., 936, 950
from partition functions, 882, 885
and rate constants, 1073
from spectroscopic data, 881
and standard free energy change, 827
and temperature, 828-30
diagram, 466
dynamic, 816
in electrolytes, 954-1000
experimental study, 831
in gases, 818
effect of inert gas, 823
in liquid systems, 822, 842-44
effect of solvent, 844
in oxidation-reduction systems, 951
metastable, 467
mobile, principle of, 831
photochemical, 1183
radioactive, 131
solvent effect in, 844
thermodynamic conditions of, 233, 234, 477, 820
and thermodynamic reversibility, 195
- Equipartition of energy, principle, 329-32, 416
- Equivalent conductance, 889
of colloidal electrolyte, 1267-69
- Ester, kinetics of hydrolysis, 1052, 1058
- Esterification equilibrium, 816, 842, 1073, 1128
- Ethane equilibrium, 1074
- Ethyl acetate, formation of, 816, 842, 1073
- Eutectic point, 747
systems, 750
with congruent melting point, 801
ternary, 801
- Eutectoid, 771
- Evaporation, *see also* Vaporization
of salt solutions, 777, 806
- Exchange energy, *see* Resonance energy
reactions, isotopic, 157-59, 176
- Excited particles, life of, 1155-56
- Exclusion principle, 80
- Exothermic reaction, 200
- Expanded surface films, 1212
- Explosion limits, 1083
- Extensive properties, 186
- Extinction coefficient, 582
- Extraction with solvent, 741-42
- Extranuclear electrons 23, 81-92
- Faraday, the, 886
and electron, 1, 13
effect, 608
laws of electrolysis, 1, 885
and ionic speeds, 912
significance of, 886
- Fatty acids, association of, 115, 319, 321, 400, 510, 680
crystal structure of, 399
structure of liquid, 510
- Fermi-Dirac statistics, 274
- Ferric chloride-water systems, 777
oxide sols, 1241
precipitation of, 1244
thiocyanate equilibrium, 844
- Ferromagnetism, 610, 611
- Ferrous sulfide, lattice type, 385-87
- Fiber structure, 403-07
- Fick's law of diffusion, 1257
- Fibrillar theory of gel structure, 1262
- First law of thermodynamics, *see* Thermodynamics
order reactions, 1046-54, 1120
consecutive, 1075
- Fission, nuclear, 173
- Flames, diffusion, 1097, 1189

- highly dilute, 1097, 1189
- Flowing junction, 943
- Flow method of equilibrium, 832
 - for reaction rates, 1049
- Fluidity, 497; *see also* Viscosity
- Fluorescence, 1155
 - of anthracene solution, 1168
 - anti-Stokes, 1155
 - and phosphorescence, 1158
 - and photochemistry, 1155
 - quenching of, 1156-57
 - resonance, 1155
 - sensitized, 1157
- Fluorides, crystal structure of, 386, 389
- Fluorite, lattice type, 388-90
- Foams, 1251
- Force constants, 569, 577-78
 - and bond lengths, 580
- Four component systems, 812
- Fourier method of crystal analysis, 362
- Fractional crystallization of solid solutions, 763
 - distillation, 717-21
 - order of reaction, 1060, 1120
- Franck-Condon principle, 574
- Free energy, 230
 - of activation, 1103, 1104, 1129
 - and affinity, 852
 - change, determination, 853
 - from E.M.F., 855, 923
 - from reaction isotherm, 853
 - from solubility, 855
 - from spectroscopic data, 880
 - from vapor pressure, 854
 - and equilibrium, 234, 819, 820
 - function, 879
 - and partition function, 879
 - standard, 827, 855
 - addition of, 856
 - and equilibrium constant, 827
 - of formation of compounds, 856
 - of ions, 1010
 - of surface, 1206
- Free radicals and chain reactions, 1082, 1084
 - and paramagnetism, 618
 - in photochemistry, 1177-79
- Free volume of liquids, 512
- Freezing mixture, 772
- Freezing point curves of two-component systems, 742-70
 - depression of solutions, 642-51
 - and activity, 689, 965
 - and molecular weight, 646
 - and osmotic pressure, 673
 - and vapor pressure, 642
 - separation of solid solutions at, 649-51, 761-69
 - and solubility, 743
 - systems with maximum 764
 - systems with minimum, 765
- Freundlich adsorption isotherm, 1196
- Friedel-Crafts reaction, 1141
- Fugacity, 301-02
 - in mixtures, 304-05, 306
- Fundamental bands, 567
 - series, atomic spectra, 40, 41, 60, 66
- Fusion, 460; *see also* Melting
 - entropy of, 462-63
 - latent heat of, 461-62
 - effect of temperature on, 462
 - mechanism of, 463
- Galvanic cells, 921; *see also* Cells
- Gamma (γ) phases in alloys, 380-81
- Gamma rays, 119, 125-26
 - diffraction of, 125
 - disintegration by, 165, 166
 - induced radioactivity by, 170
 - internal conversion of, 171
 - production by artificial disintegration, 162-64
 - production by induced radioactivity, 169, 170
 - wave length of, 125
- Gas cells, 933
 - compressibility coefficient, 310-11
 - densities, 307-14
 - by buoyancy method, 309
 - corrections in determination, 309-14
 - limiting, 307, 310-12
 - normal, 307
 - by Regnault method, 308
 - by volumeter method, 308
- ideal, *see* Ideal gas
- law, 193
 - constant, 193-94
 - for dilute solutions, 663, 671
 - for surfaces, 1214

- mean free path of molecules in, 274-76
- pressure, 248, 249-51
- reactions, homogeneous, 1048, 1057, 1063
 - catalyzed, 1129
 - on solid surfaces, 1118-26
- solid equilibrium, 778-86
- velocity of molecules in, 253
 - distribution of, 264-67
 - experimental determination, 254, 269
 - mean, 268
 - most probable, 268
- see also* Equations of state
- Gaseous ions, refraction of, 540
- surface films, 1212
- state, 244
- Gases, adsorption of, 1194
 - and critical temperature, 1197
- chemical equilibria in, 832-42
- collisions in, 276-78
- compression of, 244-46
- diffusion of, 254
- heat capacity of, 322-37
- kinetic theory, 248-85, 303
 - modifications, 285-301
- liquefaction of, 427-43
- mixtures of, 302-07
 - thermodynamics of, 304
- permanent, 427
- solubility, ideal, 702
- and solutions, analogy, 664
- thermodynamics of, 301, 304
- viscosity of, 279
 - and mean free path, 280-81
 - measurement, 283-85
 - and molecular diameter, 283
 - and temperature, 280
- Gay-Lussac's law, 183, 191-92
 - deviations from, 247
- Geiger-Müller counter, 20, 168
- Geiger-Nuttall rule, 129-30
- Gegen-ions, *see* Counter-ions
- Gelatin, sol and gel, 1255, 1256
 - swelling of, 1260, 1261
- Gelation, 1234, 1258
- Gels, 1233, 1258-66
 - elastic and non-elastic, 1258
 - imbibition by, 1259
 - precipitation in, 1264
 - structure of, 1262
 - swelling of, 1259, 1262
 - syneresis of, 1263
 - thixotropy of, 1263
- General acid and base catalysis, 1133
- Gibbs adsorption equation, 1206-08, 1213
 - application of, 1213, 1216
 - verification of, 1208, 1213
- Duhem equation, 238, 688, 707; *see also* Duhem-Margules equation
- free energy, *see* Free energy
- Helmholtz equation, 231-32, 494, 828, 924
- Glass, electrode, 995
 - formation of, 521
 - nature of, 517
 - structure of, 518-20
 - X-ray diffraction by, 518
- Glide planes, 354
- Glucose, mutarotation of, 1133, 1137
- Gold number, 1249
- Goldschmidt's rules of crystal structure, 385-91
- Goniometer, 341
- Graham's law of diffusion, 256
- Graphite, crystal structure of, 375-76
- Gronwall-LaMer treatment of strong electrolytes, 968
- Grotthuss conduction, 887, 898
 - Draper law, 1155
- Ground state, atomic, 54
- Group displacement law, 135
- Haber** process, 840
- Hair structure of, 405-06
- Halide ion catalysis, 1140
- Halogen electrodes, 947
 - electron affinity, 411
- Harmonic bands in spectra, 567
 - vibrations, 567
- Heat of activation, *see* Activation
- of adsorption, 1195, 1201, 1204
- capacity, 189
 - at absolute zero, 859, 864
 - of calorimeter, 208
 - at constant pressure and volume, 189-90, 323
 - differences, 194, 235, 236, 323, 327-28
 - of gases, 322-37

- determination, 324-25
- ratio, 323, 325-27
- and temperature, 214, 331, 335, 336
- partial molar, 243, 1010
- and partition function, 333
- and quantum theory, 332-37
- of solids, 413-21
- and temperature, 414, 420
- changes at constant pressure and volume, 181, 201
- of combustion, 203, 210
- conductivity of gases, 281
- content, 188-89
 - and heat of formation, 202
 - and Joule-Thomson effect, 286
- partial molar, 241-42, 1009
- of dilution, 242, 626
 - and Debye-Huckel theory, 968
- of dissociation from spectra, 575-76, 1165
- and energy changes, 187
- of formation, 202-03
- of fusion, 461-62
- of hydrogenation, 203, 591
- of ionization, 992
- of reaction, 200
 - determination, 207-11
 - from E.M.F., 924
 - from equilibrium constant, 830
 - and temperature, 212-15
- of solution, 241, 695
 - and lattice energy, 412
- of transition, 207, 463, 867
- units, 182
- of vaporization, 444
- Heavy water, 160, 473
- Heisenberg uncertainty principle, 18-19
- Heitler-London, covalence theory, 103
 - and hydrogen molecule, 102
- Helium, liquid, 512-14
 - ortho- and par-, 79
 - spectrum, 78
 - spectrum of ionized, 39
- Helmholtz double layer, 1220, 1227
- Hemihedral class, 346
- Hemolysis, 662
- Henderson equation, 1003, 1007
- Henry's law, 696
 - and Raoult's law, 700-01
 - deviations from, 697
- Hess's law of heat summation, 204-06
- Heterogeneous catalysis, active centers in, 1148
 - involving gases, 1144-52
 - mechanism of, 1144-47, 1149
 - in solution, 1152-54
 - chemical equilibria, 845-52
 - reactions, 1118-26, 1144-52
 - kinetics of, 1118-26
- Hildebrand hydrogen electrode, 994
 - rule, 458
- Hittorf method for transference numbers, 910, 915
- Hoar frost curve, 467
- Hofer-Moest reaction, 1038
- Hofmeister series, 1254, 1259, 1262
- Holes in liquids, 432, 500, 512
- Holohedral class, 346
- Homogeneous catalysis, involving gases, 1129-32
 - in solution, 1132-41
 - chemical equilibria, 832-45
 - reactions, 1046-1118
- Homopolar method in wave mechanics, 101
 - terms and electrovalence, 109
- Hudson's isorotation rules, 604
- Hume-Rothery rules, 380
- Hund's maximum multiplicity rule, 67, 68, 620
- Hydrates, salt, with congruent melting point, 776
 - with incongruent melting point, 773
 - vapor pressure of, 781-83
- Hydration and transference number, 920-21
 - in solution, 683
 - of ions, 412, 621, 920
 - of lyophilic sols, 1253
- Hydrocarbon chains, cross section, 399, 1210, 1212
 - crystal structure, aliphatic, 400-01
 - aromatic, 403
 - structure of liquid, 508
- Hydrogen atom, mass of, 264
 - spectrum of, 37-39
 - Bohr theory, 45-48
 - wave mechanics of, 77

- bond, 113-16
 in carboxylic acids, 115, 510
 in crystals, 393-94
 energy of, 114
 in liquid mixtures, 728
 in liquids, 509-10
 in vitreous state, 521
 in water, 114
-bromine reaction, photochemical, 1171
 thermal, 1079
carrier, 1040
cell, E.M.F., 933
-chlorine reaction, 1080-81
 photochemical, 1173
-deuterium reactions, 1061
electrode, 994
 standard, 941
entropy of, 876
fluoride, association of, 318
 as solvent, 976
 structure of, 115
heat capacity of, 335-36
iodide decomposition, photochemical, 1170
 thermal, 1072, 1096
-iodine reaction, 831, 1072, 1080
ion, catalysis by, 1110
 concentration, determination of, 993-95, 1001
 by electrodes, 994-95
 by indicators, 587, 1001-02
 during neutralization, 1002-05
 in acid-salt mixtures, 1003, 1007
 conductance of, 898
 in solution, 898, 975
isotopes, discovery of, 147, 172
 separation of, 156-57, 158-59
 spectrum of, 147
 see also Deuterium, and Tritium
molecule, wave mechanics of, 102-03
 -ion, 106
ortho- and para-forms, 94-97
 conversion, 1060, 1082, 1093
overvoltage, 1017-22
-oxygen cell, 946
 gas reaction, photosensitized, 1175, 1181
 thermal, 1082
 peroxide, decomposition, 1054
 catalytic, 1140, 1153
 rotation in solid, 423
 scale of potentials, 940
Hydrogenation, heat of, 203, 591
Hydrogen-like atoms, spectra of, 40
 theory of, 45-50
Hydrolysis constant, 985-88
 determination of, 989-90
 degree of, 986
 and strength of acid and base, 984.
 of esters, 1052, 1058
 of salts, 983-90
 of strong acid and weak base, 986
 of weak acid and strong base, 984
 of weak acid and weak base, 987
Hydronium or Hydroxonium ion, 975
Hydroxyl ions, catalysis by, 1054, 1132-40
Hydroxylic compounds, association, 115, 509-10, 679-82
 resonance in, 115
Hypsochromic effect, 583

Ice, forms of, 472-73
 structure of, 115, 394
Ideal gas, 191-95
 adiabatic expansion, 199
 entropy of, 224
 equation of state of, 193
 internal energy of, 191
 reversible work of expansion, 196-97
 thermodynamics of, 191-200
 solution, 626, 675; *see also* Solution
Imbibition by gels, 1259
Immiscible liquids, 733
 vapor pressure of, 734
Incomplete miscibility of liquids, 722
 of solids, 766-69
Incongruent melting point, 804
Incongruently saturated solution, 809
Independent migration of ions, 894
Indicators, adsorption, 1248
 isotopic, 175-78
 neutralization (*or* acid-base), 999-1002, 1005
 constants, 1000, 1001
 determination of pH by, 1001
 equilibria, 999
 range of, 1000

- in titrations, 1005
 - oxidation-reduction, 952, 1009
 - radioactive, 175-78
 - Induced anisotropy, 603
 - dissymmetry, 602
 - oxidation, 1143
 - polarization, 549, 554, 555
 - radioactivity, *see* Radioactivity
 - Induction effect, in molecular attraction, 299
 - period of, 1077, 1173
 - Infrared spectra, 565
 - of liquids, 580
 - Inhibition of reaction, 1142
 - Inner quantum number, 62, 334
 - Integral heat of dilution, 242, 968
 - of solution, 242
 - Integrated reflection of X-rays, 361
 - Intensity, radiation, 581, 1162, 1169
 - Intensive properties, 186
 - Interatomic distances, *see* Atomic distances
 - Interfacial tension, 481, 484
 - and adsorption, 1215
 - and emulsification, 1279
 - measurement of, 490-91
 - Interionic attraction theory, 903, 956
 - Intermediate compound theory of catalysis, 1131, 1150
 - Internal conversion, 171
 - energy, 186, 329, *see also* Energy
 - heat capacity due to, 329
 - and partition function, 333
 - pressure, 289, 479-80
 - and abnormal molecular weights, 683
 - and polarity, 677
 - and solubility, 702, 722, 744
 - Interstitial compounds, 381-82
 - structures, 382
 - Intrinsic energy, *see* Internal energy
 - Ion association, 969
 - conductance, 894
 - Ionic atmosphere, 903, 957
 - of colloids, 1247
 - radius of, 958
 - attraction in crystal lattice, 407-09
 - bonds, *see* Electrovalence
 - conductance, 894
 - equilibria in water, 982
 - mobility, 909
 - micelles, 1267
 - product, 978
 - of water, determination of, 990-91
 - effect of ions, 993
 - effect of temperature, 992
 - radii, ratio of, 382-83
 - and lattice type, 385-91
 - reactions, 1114-18
 - repulsion in crystal lattice, 407-09
 - strength, 956
 - and activity coefficients, 956, 959, 960
 - and reaction rates, 1116, 1138
 - and solubility, 963
 - terms in covalence, 101
 - velocities, 895, 908-10
- Ionization constant, *see* Dissociation constant
- potentials, 50
 - calculation, 53
 - measurement, 50-53
- Ion-pairs, 907, 969
- triplets, 897, 969
- Ions complex, *see* Complex ions
- conductance of, 894
 - and viscosity, 897
 - application of, 898
 - definition of, 884
 - deformation of, 538-40
 - effective diameter, 967
 - independent migration of, 894
 - magnetic moments of, 619
 - susceptibility, 614, 617, 619
 - mobility, 909
 - of oxyacids, structure of, 391
 - polarizability, 538-41
 - refraction of, 539-41
 - solvation of, 920
 - speed of, 895, 908-10
 - standard entropies of, 1011-13
 - free energies of, 1010
- Iron-group metals, electrodeposition, 1023
- steam equilibrium, 850
- Irreversible electrode phenomena, 1013-40
- thermodynamic processes, 207-09
 - entropy change in, 227-29
- Isentropic change, 225
- Isobares, 153

- Isobaric isotopes, 153, *see also* Isomeric nuclei
- Isoelectric point, of ampholytes, 998
of proteins, 1255
- Isolated reactions, 1045
- Isomeric nuclei, 171-72
- Isomorphism, 346-50
applications of, 349
definition, 348
and ionic radii, 392-93
law of, 347
and solid solutions, 349
tests for, 348-50
- Isopiestic solutions, 632
- Isopolymorphism, 350
- Isorotation rules, 604
- Isoteniscope, 447
- Isotherm, *see* Reaction isotherm
- Isothermal change, 219-20
evaporation, 777, 806
invariant point, 804
- Isotonic solutions, 632, 661
- Isotopes, 30, 135
atomic weights of, 137
detection by atomic spectra, 147
by mass spectra, 142-47
by molecular spectra, 148
distribution in nature, 151
effect in spectra, 147, 148, 567, 572
equilibria of, 157, 158-59
exchange reactions of, 157-59, 176
as indicators, 175-78, 1074, 1142
isobaric, 153, 171-72
and mass spectra, 142-47
nonradioactive, 137-38, 149
nuclear structure of, 136-37
and optical spectra, 147, 148
and periodic classification, 135, 151
radioactive, 135-37
reactions of, 1105
separation by chemical methods, 157-59
by diffusion, 153-54
by distillation and evaporation, 155-56
by electrolysis, 156-57
factor of, 160
by positive rays, 156
by thermal diffusion, 154-55
- Isotropic substance, 346
- jj- coupling, 68
- Joule, the, 182, 184, 922
- Thomson coefficient, 287, 293
calculation of, 292-93
equation for, 287
effect, 190, 286
cooling by, 441
inversion temperature, 293-94
- K electron capture, 170
- Kelvin temperature scale, 222
- Keratin, structure of, 406
- Kerr constant, 557, 561, 562
and molecular structure, 560-63
and polarization, 559-63
sign of, 559
effect, 550, 557-63
in liquids, 562
- Ketones, photolysis of, 1177
- Kilogram, standard, 181
- Kinetic activation, theory of, 1090
activity factor, 826, 1115
energy and temperature, 252
of a molecule, 252-53, 303; *see also* Translational energy
theory of gases, *see* Gases
theory of vapor pressure, 443
- Kinetics, chemical, 1044-1190; *see also* Reactions
of heterogeneous reactions, 1118-26
- Kirchhoff equation, 212-15, 453, 462, 861
- Kohlrausch law of migration of ions, 894
regulating function, 914
- Kolbe electrosynthesis, 1037
- Konowaloff's rule, 714
- Kopp's law of heat capacity, 415
- Kossel, valence theory, 97
- Lambert's law, 582
- Lambda (λ) point, 512
- Landé splitting factor, 65, 617
- Landolt reaction, 1078
- Landsberger boiling method, 639
- Langmuir, adsorption isotherm, 1198
atomic structure theory, 81
surface reaction theory, 1118-23
valence theory, 98-99
- Latent heat, *see* Fusion, Sublimation, Transition, Vaporization

- image, 1187
- Lattice, compounds, 397
 - crystal, *see* Space lattice
 - energy, 407-13
 - and heat of solution, 412
 - and proton affinity, 412-13
 - and valence type, 412
 - layer, 376, 389
- Laue photograph, 371
- Layer lattices, 376, 389
 - lines, in rotation photographs, 372
- Lead, atomic weight of isotopes, 138
 - isotopic composition, 151
- Le Chatelier principle, 831
- Lewis, acid-base definition, 974
 - valence theory, 98
- Lewis-Randall ionic strength principle, 956
- Libration in solid state, 422
- Liesegang rings, 1264
- Limiting current, 1027
 - density, 307
 - pressure method for gas density, 312
- Linear polarization, 597
- Linkages, characteristic frequencies of, 579
 - force constants of, 578
 - heat of formation of, 588-90
- Liquefaction, 427-43
- Liquid crystals, 514-17
 - cholesteric, 516
 - nematic, 515
 - smectic, 515
 - theory of, 516
 - junction potentials, 931, 943
 - junctions, types of, 943
 - liquid systems, distribution in, 735-42
 - mixtures, boiling points, 715-17
 - distillation of, 717-21
 - vapor pressures, 704-715
 - solutions, *see* Solutions
 - state, 479-514
 - theory of, 512
- Liquids, chemical equilibria in, 825, 842
 - fluidity, *see* Viscosity
 - inmiscible, 733
 - internal pressure of, 479-80
 - partially miscible, 721-33
 - spectra of, 580-88, 1167
 - structure of, 506-10
 - theory of, 510
 - surface tension of, 481-96; *see also* Surface Tension
 - vapor pressure of, 704-715; *see also* Vapor pressure
 - viscosity of, 496-504
 - X-ray diffraction by, 504-10
- Liquidus curve, 747, 761
- Lone-pair of electrons, 100, 538
- Long-chain molecules, crystal structure, 399-402
 - surface films of, 1210-15
 - viscosity of solutions, 500, 503
- Lorentz-Lorenz equation, 529
- Loschmidt number, 282
- Lowry, acid-base definition, 974-75
- Low temperature, by adiabatic demagnetization, 442
 - by adiabatic expansion, 441
 - by external work, 441
 - by Joule-Thomson effect, 441
- Luminescence, 1188
- LS-coupling, *see* Russell-Saunders coupling
- Lyman series, 49-50, 54
 - of deuterium, 147
- Lyolysis, 984; *see also* Hydrolysis
- Lyophilic systems, action on lyophobic, 1249
 - characteristics of, 1235
 - electrophoresis in, 1252
 - electroviscous effect in, 1251
 - hydration of, 1253
 - organosols, 1270
 - precipitation by alcohol, 1253, 1255
 - salting out of, 1254
- Lyophobic systems, characteristics of, 1235
 - electrophoresis in, 1240
 - mutual action of, 1248
 - optical properties of, 1235
 - origin of charge, 1240
 - precipitation by electrolytes, 1243
 - protection of, 1249
 - shape and color of particles, 1238
 - size of particles, 1237
- MacLeod's equation, 494, 526
- Macromolecule, 376

- Madelung's constant**, 409
- Magnetic moment**, 65, 616-17
- permeability, 610
 - properties of electron, 64
 - of rare-earth ions, 91-92
 - quantum number, 63
 - rotation, 608-09
 - susceptibility, 610
- Magnetism**, *see* Diamagnetism, and Paramagnetism
- Magneton**, Bohr, 65, 616
- Weiss, 616
- Marriotte's law**, *see* Boyle's law
- Mass action, law of**, 816, 845; *see also* Equilibrium
- Mass of atom**, 264
- and energy relationship, 7
 - in atomic disintegration, 165-66
 - and atomic weights, 166-67
 - number, 150
 - spectrograph, 142-47
 - Aston's, 142
 - Bainbridge's, 144-45
 - Dempster's, 144
 - double focusing, 146
 - Nier's, 146-47
 - spectrum, 143, 146
- Matter waves**, 15, 73
- states of, 244
- Maxwell-Boltzmann law**, 95, 270-72
- statistics, 270
- Maxwell's distribution law**, 264-67
- experimental verification, 269
 - simplified form of, 269-70
 - refractive index relationship, 536
- Maximum boiling point system**, 720
- bubble pressure, 488-89
 - freezing point system, 764
 - work of expansion, 198
 - function, 230
 - in isothermal change, 197-98, 230
- Mean free path**, 274-77
- in mixtures, 276
 - and viscosity, 280-81
- Mechanical equivalent of heat**, 182, 183-84
- Melting-point curve**, 460; *see also* Fusion
- effect of pressure on, 461, 466, 472, 513
 - of small particles, 496
 - systems with congruent, 755
 - with incongruent, 759
 - with maximum, 764
 - with minimum, 765
- Membrane equilibrium**, 1259-62
- and osmosis, 1261
 - and proteins, 1261
 - and swelling of gels, 1261
- Menschutkin**, solvent effect on reaction rate, 1113
- Menzies**, vapor pressure method, 629; *see also* Smith and Menzies
- Mercury dropping cathode**, 1027
- Mesomorphic state**, *see* Liquid crystals
- Meson**, 30
- Mesophases**, 516
- Metallic alloys**, 379-81
- bond, 378-79
 - state, 377-79
- Metal sols**, 1242
- Metals**, corrosion of, 1033-35
- electrodeposition of, 1022-26
 - mobility of electrons in, 379
 - order-disorder transformations in, 382
 - physical properties of, 379
 - separation by electrolysis, 1024
 - simultaneous deposition, 1025
 - structure of, 377-78
- Metastable forms**, 469-71
- equilibrium, 467
- Meter**, standard, 181
- Mho**, 889
- Micelles**, 1241, 1267
- neutral and ionic, 1268
- Microcrystalline structure of liquids**, 511
- Migration of ions**, independent, 894
- Miller indices of crystals**, 344
- Minimum boiling point systems**, 718
- conductivity, 896, 969
 - freezing point systems, 765
- Miscibility gap**, 723
- Mitscherlich's law of isomorphism**, 347
- Mix-crystals**, 349; *see also* Solid solutions
- Mixed crystals**, *see* Mix-crystals
- Mixtures of gases**, *see* Gases
- Mobile equilibrium**, principle of, 831
- Mobilities of colloidal particles**, 1239
- of ions, 909

- Molal depression constant, 645
 elevation constant, 636
Molality defined, 636
Molar absorption coefficient, 582
 conductance, 889, 901
 dispersivity, 534
 extinction coefficient, 581
 heat capacity, *see* Heat capacity
 magnetic rotation, 608
 polarization, 535; *see also* Polarization
 refraction, 528-32; *see also* Refraction
 rotation, 596
 volume, 524, 526, 529, 536
 additivity of, 525
Molecular association, 113-15, 318, 509, 679-82
 of alcohols, 115, 509
 of carboxylic acids, 115, 319, 321, 400, 510, 680
 of *p*-cresol, 681
 and distribution ratio, 737
 and hydrogen bonds, 113-15
 and molecular weights, 679-82
 in solution, 677
 and surface tension, 493
attraction, 286, 289
 origin of, 298-301
compounds, 100, 116, 397
crystal lattice, 377, 381
diameter from diffusion, 281, 285
 from mean free path, 282
 from polarization, 542
 from van der Waals constant, 294
 from viscosity, 283-84
orbitals, 101
 and valence, 105-06
orientation at surfaces, 1209-12
polarizability, 534
repulsion, 300, 409
spectra, *see* Spectra
velocity distribution, Maxwell's law, 264-69
weights, and gas density, 307-18
 correction for nonideality, 309-14
 by ultracentrifuge, 1256-58
 in solution, 627, 637, 646, 672
 abnormal, 678-83
 review of, 673-74
 significance of, 670, 676
 by viscosity, 501
 of vapors, 314-18
Molecularity of reaction, 1044
Molecules, polar and monopolar, 108-09
 velocity of, 249-51
 determination, 254
 mean, 268
 most probable, 268
 root mean square, 253
Mole fraction, 303
Moments of inertia, 569
Monatomic gas, entropy, 873
 molar heat capacity, 331, 334
Monodisperse systems, 1257
Monolayer, 1211; *see also* Surface films
Monotropy, 470-71
 and enantiotropy, 471
Morse equation, 573
Moseley's law, 25, 70, 82
Mosotti-Clausius equation, 534-35, 543
 derivation of, 535
Moving boundary method for transference numbers, 913, 916
Multimolecular layers, adsorption, 1202
 built up, 1214
Multiplet levels, atomic, 41, 62, 69
 molecular, 334
 spectra, *see* Spectra
N-chloracetanilide, isomeric change, 177, 1053
Negative adsorption, 1205, 1216
 catalysis, 1126, 1142, 1143
Negatron, 21
Nematic phases, 515
Neptunium, 173
Nernst approximation formula, 862
 equation, 948
 heat theorem, 857-64
 application to gas reactions, 860
 solution pressure theory, 947
 -Thomson rule, 896
Net-planes, 352
Net work, 230
Neutral salt error of indicators, 1002
Neutralization curves, 1002-05

- and hydrolysis, 983
- indicators, 999–1002, 1005
- titration and *pH*, potentiometric, 1008
- Neutretto, 30
- Neutron, 28–31
 - as constituent of nucleus, 29–30, 137, 174
 - disintegration by, 162
 - mass of, 29, 166
 - and meson, 30
 - and nuclear fission, 173
 - production in disintegration, 162–65
 - production of induced radioactivity by, 169–70, 173–74
- Nitrate ion, structure of, 112, 392
- Nitric oxide, formation equilibrium, 834
 - reactions, 1063
- Nitrogen, active, 1064, 1189
 - pentoxide, decomposition of, 1048–49
 - in solution, 1113
 - tetroxide, dissociation of, 321, 836
 - in solution, 843
- Nonaqueous media, colloids in, 1270
 - conductance in, 896
 - E.M.F. in, 953
 - zeta-potential in, 1228
- Nonbonded radii, 384
- Nonideal solutions, 676–83
 - thermodynamics of, 683–90
 - systems, equilibria in, 825–26
- Nonpolar molecules, 108–09
- Nonstationary chains, 1083
- Nuclear charge, 25
 - energy levels, 171
 - fission, 173
 - isomerism, 153, 171–72
 - spin, 92–97
 - and entropy, 875
 - and ortho- and para- states, 94–97
 - and partition function, 875
 - quantum number, 93–94
 - stability, 174–75
 - theory of atom, *see* Atomic structure
- Nuclei, constituents of, 30
 - stability of, 174
 - unstable, 167–70
- Nucleus, *see* Atomic nucleus
- Occlusion, 784, 1194, 1204
- Odd-electron bonds, 106–07
 - and magnetism, 618
- Ohm, 889
 - reciprocal, 889
- Ohm's law, 888
- Onsager equation, 904–07
- Opposing reactions, 815, 1069–75
 - and equilibrium constant, 1073
 - kinetics of, 1069
- Optical activity, 596–608
 - anisotropy, induced, 603
 - electron, 42, 66
 - exaltation, 530, 534
 - rotation, 596
 - and absorption bands, 601–02
 - and chemical structure, 603–04
 - and chromophoric centers, 602
 - and circular dichroism, 605
 - and circular polarization, 598
 - dispersion of, 598–601
 - isorotation rules, 604
 - molar, 596
 - normal and anomalous, 600
 - simple and complex, 599–600
 - solvent effect in, 605
 - superposition rules, 603
 - theories of, 607
 - superposition, 603
- Optic axis of crystal, 346
- Orbitals, atomic, 101
 - molecular, 101, 102; *see also* Wave functions
- Order-disorder in liquid, 511
 - in melting, 463
 - transformations, 381
- Order of reaction, 1045–69; *see also* Reaction
- Organic compounds, crystal structure, 399–407
 - oxidation-reduction systems, 951
 - see also* Nonaqueous
- Organosols, 1270
- Orientation effect in molecular at
 - traction, 298
 - polarization, 545
 - at surfaces, 1209–12
 - in catalysis, 1149
- Orthobaric densities, 431–32
- Orthohelium, spectrum of, 79
 - structure of, 81

- Ortho-hydrogen, 94-97
 heat capacity, 335-36
 ratio to para-, 95-96
Ortho-states, 94
Oscillator, energy of, 32, 567
Osmosis, 651
Osmotic coefficient, 685-86
 and activity coefficient, 686
 pressure, 652-74
 comparison, 660-62
 definition of, 652-53
 and depression of freezing point, 672
 effect of concentration, 654, 663, 670
 effect of temperature, 654, 663
 and elevation of boiling point, 672
 measurement of, 653, 657-60
 molecular weight from, 672
 specific, 673
 theories, 663-68
 properties, 625, 672
Ostwald's dilution law, 955
Overgrowths, crystal, 349
Overpotential, *or* Overvoltage, 1017
 bubble, 1018-19
 and corrosion, 1033
 at definite c.d., 1019
 and electrolytic reduction, 1039
 hydrogen, 1017-22
 oxygen, 1032-33
 theories of, 1020-22, 1032
Oxidation, electrolytic, 1035, 1040
 of anions, 1036
 reversible, 1035
 induced, 1143
 potentials, 939
 reactions and luminescence, 1189
 -reduction electrodes, 949-50
 equilibrium, 951
 indicators, 952, 1009
 systems, 939
 with organic compounds, 951
 titrations, 1008
Oxide electrodes, 994
Oxides, crystal structure of, 389
 dissociation of, 780, 847, 864
Oxonium ions, 958, 975, 976
Oxyacids, ions, crystal structure, 391
Oxygen carriers, 1040
 cell, *E.M.F.* of, 933
 electrode, 946, 994
 evolution, 1032
 -hydrogen cell, 946
 isotopes, 148, 150, 157, 158
 molecular structure of, 107
 paramagnetism of, 107, 618
Ozone, photosensitized decomposition, 1182

Packing fraction, 152-53
Palladium-hydrogen system, 784-86
Parachor, 526
 anomalies, 528
 applications of, 527
Paraffins, crystal structure of, 400-01
 structure of liquid, 508
Para-hydrogen, 94-97
 heat capacity, 335-36
 ratio to ortho-, 95-96
Parallel growths, crystal, 349
Paramagnetic elements, 107, 618
 susceptibility, 65, 616
 of complex ions, 620
 of free radicals, 618
 of ions, 89, 619
 of rare-earth ions, 91-92
Paramagnetism, 610, 611, 615-22
 and electron spin, 617
Parameter, atomic, in crystal, 360
Para-states, 94
Parhelium, spectrum of, 79
 structure of, 81
Parkes's process, 795
Partial molar, free energy, 238; *see*
 also Chemical potentials
 heat capacity, 243
 relative, 243, 1010
 heat content, 242, 1009
 relative, 242, 687, 1010
 quantities, 237-43
 determination of, 239-41
 relative, 242-43
 volume, 239
 pressure, 302-04
 and equilibrium constant, 819-20
 and gas solubility, 698
 law of, 302
Partially miscible liquids, 721-33,
 792-800
 distribution between, 735

- influence of added substances, 729
 - and solid phases, 753
 - with three components, 796
 - vapor pressure of, 730
- Particle in a box, 872
- Partition function, 333
 - and entropy, 870-72
 - and equilibrium constant, 880-82
 - and free energy, 879
 - and heat capacity, 333
 - for internal energy, 333
 - in reaction rates, 1100, 1101, 1145
 - for rotation, 334-36, 881
 - for translation, 333, 872
 - for vibration, 337
- law, *see* Distribution law
- Paschen-Back effect, 64
- Paschen series, 38, 49-50
- Passivity, 1028-32
 - chemical, 1029
 - mechanical, 1031
 - theories of, 1030
- Pauli exclusion principle, 80
- P-branch, 598
- Pectization, 1233
- Peptization, 1242, 1274
- Perfect gas, *see* Ideal gas
- Periodic reaction, 1078
- Period of induction, 1077, 1173
- Peritectic point, 768
- Perpetual motion of second kind, 217
- Perrin's work on Brownian movement, 255-63
- Persorption, 1204
- Pfeffer's measurement of osmotic pressure, 643
- Pfund series, 39, 49-50
- pH, 982; *see also* Hydrogen ion concentration
- Phase, definition, 474
 - diagram, 466
 - for liquid helium, 513
 - for sublimation, 468
 - for sulfur, 471
 - for water, 465-66, 473
 - rule, 473-78
 - applied to one-component systems, 476
 - to gas-solid systems, 778-86
 - to solid-liquid-vapor systems, 786
 - to solubility of gas in liquid, 693
 - to two-component systems, liquid, 723, 731
 - with solid phases, 745-78
 - thermodynamic derivation, 477
 - volume theory of emulsions, 1276
- Phenols, association of, 115
- Phosphorescence, 1158
- Photochemical decomposition, *see* Photolysis
- Photolysis
 - dimerization of anthracene, 1168, 1184
 - equilibrium, 1183-84
 - equivalence law, 1159
 - experimental methods, 1160
 - hydrogen-bromine reaction, 1171
 - hydrogen-chlorine reaction, 1173-76
 - isomeric change, 1187
 - predissociation reaction, 1167
 - quantum efficiency (or yield), 1160
 - reaction, influence of temperature, 1185
 - in liquids, 1179
 - primary, 1164
 - secondary, 1168
 - in solids, 1186-88
 - silver halide decomposition, 1187
- Photochemistry, 1154-90
 - and spectra, 1164-67
- Photoelectric effect, 4, 33-34
- Photographic emulsion, 1186, 1275
 - plate, 1187
 - desensitization and sensitization, 1187
- Photolysis, 1170
 - of aldehydes, 1176, 1178
 - of ammonia, 1176
 - of hydrogen iodide, 1170
 - of ketones, 1177
- Photon, 14, 33
- Photosensitization, 1180
- Photosensitized reactions, 1157, 1176, 1180
 - with chlorine, 1176, 1180, 1182
 - with chlorophyll, 1183
 - with mercury, 1181
- Photostationary state, 1183

- Photosynthesis, of carbohydrates, 1183
- Pitchblende, 118, 174
- Plait point, of binary system, 793
- Planck's constant, 32
quantum theory, 31-33
- Plasmolysis, 661
- Platinum sol, 1242
as catalyst, 1153
- Plutonium, 173-74
- Poggendorff compensation method, 924
- Point groups, crystallographic, 344
- Poise, 279, 499
- Poisoning of catalyst, 1148, 1151, 1153
- Poisson's equation, 957
- Polarity of bond, 108
and internal pressure, 677
- Polarizability, 534-36
atom and electron, 536
and crystal structure, 386-89
of electron groups, 538
of ions, 386, 389, 538-41
and Kerr effect, 558-63
and molecular radius, 541
- Polarization, 534-43
atom, 546
electrolytic, *see* Electrolytic polarization
electron, 536
ellipsoid of, 557, 560
induced, 535
of light, 597
molar, 535
orientation, 545
and temperature, 546
- Polar linkages, 108
molecules, 107-09, 543
solvents, 586, 677, 703, 729
- Polarograph, 1027
- Polonium, discovery of, 118
radioactivity of, 133
- Polyatomic molecules, entropy of, 874
- Polydisperse systems, 1257
- Polymorphic changes, 463
enantiotropic, 469
monotropic, 469
- Polymorphism, 350, 468
- Positive rays, 2
analysis by, 138-47
and isotopes, 141-47
separation of isotopes by, 156
- Positron, 20-22
destruction of, 22
mass and charge of, 21
production in induced radioactivity, 167-70
- Potential, critical, atomic, 50
electrode, 937
sign of, 940
standard, 937, 944-47, 949
energy curves, 103, 573-75
field distribution in, 272
ionization, 50
liquid junction, 931
resonance, 50
- Potentiometer, 925
-voltmeter method, 926
- Potentiometric titration, 1008
- Powder method of crystal analysis, 373
- Precipitation titration, 900, 1009, 1248
conductometric, 900
potentiometric, 1009
- Predissociation decomposition, 1167
spectra, 1166
- Pressure, *see* Cohesion, Gas, Internal, Partial, Reduced, etc.
- Primary salt effects, 1115, 1138
- Primitive translations, 352
- Principal quantum number, 56
apparent, 59
series of spectra, 40, 41, 60, 66
- Probability and entropy, 229
factor in reaction velocity, 1098
significance of, 1102
- Promoters, in catalysis, 1152
- Protective action, 1249
- Proteins and membrane equilibrium, 1261
molecular weights of, 1258
sols, isoelectric point of, 1255
viscosity of, 1255
swelling of, 1262
structure of fibrous, 405-06
- Protium, 159
- Protogenic solvent, 975, 976
- Proton, 19, 30
affinity of ammonia, 412-13
affinity of water, 413
as constituent of nuclei, 30, 137, 174
disintegration by, 163-64

- and induced radioactivity, 168-69
- and meson, 30
- production in disintegration, 160-64
- product of induced radioactivity, 168-70
- transfer in acid-base equilibria, 975-77
 - in catalysis, 1138
- Protophilic solvent, 975
- Pseudo-bimolecular reaction, 1059
 - binary system, 752
 - ideal solution, 678
 - unimolecular reaction, 1051
- Purple of Cassius, 1250
- Pyrite, lattice type, 388-90
- Q-branch, 569, 571, 577
- Quadruple point, 786
- Quanta, 32
- Quantization of direction, 64
- Quantum defects, 59
 - efficiency or yield, 1160
 - data of, 1163
 - determination of, 1161
 - high, 1163, 1173
 - in hydrogen-chlorine reaction, 1173
 - in liquids, 1179
 - low, 1163, 1172
 - in photolysis of aldehydes and ketones, 1178
 - in photosynthesis, 1183
 - mechanics, *see* Wave mechanics
 - number, 44, 47, 55-63, 67, 83-92
 - apparent, 59
 - azimuthal, 55-58
 - inner, 62
 - magnetic, 63
 - principal, 56
 - radial, 56
 - resultant, 67
 - spin, 61-62
 - and wave mechanics, 77
 - statistics, 273-74
 - theory, 31-34
 - of heat capacity, gases, 332-37
 - solids, 417-21
 - of spectra, 43, 563
- Quenching of fluorescence, 1156
- radius, 1157
- Quinhydrone electrode, 995
- Quinone-hydroquinone systems, 952, 995
- Radial distribution method, for gases, 594
 - for glasses, 519
 - for liquids, 505
- quantum number, 56
- Radiation, quantum theory of, 31, 43
- theory of chemical reaction, 1089, 1107
- Radical, chain reactions, 1084; *see also* Free radicals
- Radii of atoms and ions, *see* Atomic radii, and Ionic radii
- Radioactive, average life, 128
- constants, 128-29
- decay, 126
- disintegration, rate of, 128
 - branched, 133
- emanations, 126-27
- end-product, 127, 137
- equilibrium, 131
- group displacement law, 135, 136-37
- half life, 129, 133-34
- radiations, 118
- series, 127, 132
- tracers, 175-77, 1074, 1142, 1218
- Radioactivity, 118-38
 - discovery, 118
 - induced (or artificial), 167-74
 - by α -particles, 167, 168
 - by γ -rays, 170
 - by deuterons, 169
 - internal conversion in, 171
 - by neutrons, 169-70
 - by protons, 168-69
 - theory of, 127
 - and wave mechanics, 130
- Radioelements, and periodic table, 134-35
 - separation of, 132
- Radium, discovery of, 118
- radioactivity of, 129, 131, 133
- Radon, 127
- Raman spectra, 576
- Ramsay-Young vapor pressure rule, 455
- Raoult's law, 627, 635

- deviations from, 676-78, 683, 711-13, 722
- and Henry's law, 700-01
- and ideal solutions, 675
- and liquid mixtures, 709
- test of, 626
- theoretical basis of, 675-76
- Rare-earth elements, atomic numbers, 27
- structure, 90
- ions, magnetic moments, 91-92
- Rast's micro-method for molecular weights, 648
- Rational activity coefficient, 960
- indices, law of, 343
- R*-branch, 568
- Reactions, autocatalytic, 1061, 1078
- bimolecular, 1057, 1060, 1112
- chain, 1080-85
- complex, 1065
- consecutive, 1075-85
- first order, 1046-51, 1085, 1120
- fractional order, 1060, 1120
- gas phase and solution compared, 1111-13
- heterogeneous, 1118-26, 1144-54
- high order, 1065
- homogeneous, 1046-1118, 1129-44
- ionic, 1114-18
- isochore, *see* van't Hoff equation
- isolated, 1045
- isotherm, 823-25
- kinetics of, 1044-1180
- mechanism by radioactive indicators, 177
- opposing, 1069-75
- order of, 1045
- determination of, 1066-69
- periodic, 1078
- photochemical, *see* Photochemistry
- photosensitized, 1180
- pseudo-bimolecular, 1059
- pseudo-unimolecular, 1051
- rate of, 1044
- and activity, 826
- and adsorption, 1124
- and concentration, 817, 826, 1044
- and free energy of activation, 1104
- solvent effect, 1113
- and temperature, 1087
- theory of, 1089-1111
- retardation of, 1121, 1123
- salt effects in, 1115
- second order, 1054-60
- side, 1085-87
- simultaneous, 1069-87
- slow, 1098, 1104
- specific rate of, 1046
- third order, 1062-64
- velocity constant, 1046
- zero order, 1045, 1122
- Recalescence, 464
- Reciprocal salt pair, 812
- Rectilinear diameter, law of, 431
- Redox, *see* Oxidation-Reduction electrodes
- Reduced equation of state, 439
- mass, of atom, 47
- of molecule, 148, 566
- pressure, temperature and volume, 439
- Reduction, electrolytic, 1035, 1039-40
- reversible, 1035
- potential, 940
- Reference electrode, 941
- Refraction, 528
- additivity, 529
- anomalies, 530
- of bonds, 537
- of electron groups, 538
- at infinite wave length, 529
- of ions, 538-41
- of mixtures, 531
- molar, 528-32
- and polarization, 536
- and molecular radius, 549
- volume, 529
- specific, 529
- Refractive dispersion, 533
- anomalies, 533-34
- index, 528, 532
- and density, 528
- and dielectric constant, 536
- and dispersion, 533
- measurement, 532
- and optical rotation, 598
- Refractometer, 532
- Repulsion forces, 288, 300
- in crystals, 407-09
- Relaxation effect, 904
- time of ions, 904

- Resistance, specific, 889
Resistivity, 889
Resonance, 79–80, 111–13
 in acids, amides and esters, 113
 in benzene, 112, 590
 and bond energy, 590
 in carbon dioxide, 112
 and color, 584–85
 and dipole moment, 556
 energy, 80, 102–03, 112–13, 1093
 fluorescence, 1155
 in graphite, 375
 and helium spectrum, 79
 and hydrogen molecule, 102–03
 and interatomic distances, 376, 403, 595
 and molecular structure, 111–13
 lines in spectra, 50
 in oxyacid ions, 112
 potentials, *see* Critical potentials
Retardation of homogeneous reactions, 1142
 of surface reactions, 1121, 1123, 1151
Reticular density, 352
Retrograde solubility, 795
Reversible cells, 216, 921; *see also* Cells
 electrodes, 922
 processes, thermodynamically, 195–200, 219–22
 condensation and vaporization, 196
 cyclic, 219–22
 efficiency of, 222
 expansion, 196
 reactions, 815, 1069–75
Ritz combination principle, 38–40, 45, 65
Rotating crystal method, 372
Rotation of diatomic molecule, 330
 restricted internal, 877
 in solid hydrogen, 423
 in solid state, 422–24, 463
 see also Magnetic rotation, and Optical rotation
Rotational energy, levels, 95–96, 336, 564, 566
 of molecules, 330, 566
 entropy, 875
 heat capacity, 334–36
 partition function, 334–36
 spectra, 565, 566
Rotatory dispersion, 598–601
 normal and anomalous, 600–601
 simple and complex, 599
Row lines, in rotation photographs, 372
Rubber, structure of, 406–07
Russell-Saunders coupling, 67–68
Rutherford and Soddy, theory of radioactive disintegration, 127
Rutile, lattice type, 388–90
Rydberg number, 37, 39, 40, 47, 48, 70, 147

Sackur-Tetrode equation, 874
Saponification of esters, 1058
 of dibasic acids, 1076
Salt bridge, 944
Salt effects in reaction velocity, primary, 1115–18, 1138
 secondary, 1139–40
Salting-out effect, in binary liquid systems, 729
 in Debye-Hückel equation, 967
 with gases, 700
 with lyophilic sols, 1254
Salt-water systems, 771–78, 786–90, 802–13
 solubilities in, 773–77
 three component, 803
 vapor pressures of, 781
Salts, acid, 812
 basic, 812
Saturated solutions, vapor pressure of, 789
Saturation vapor pressure, *see* Vapor pressure
Schreinemaker's residue method, 802
Schrödinger equation, 75, 872
Schulze-Hardy rule, 1243
Screening constant, 72
Screw axes, 354
Secondary salt-effects, 1139–40
Second law of thermodynamics, *see* Thermodynamics
 order reactions, 1054–60
Sedimentation constant, 1256
Seeding, 647, 749
Selection principle, 58
Sellmeier equation, 533, 599
Semipermeable membranes, 651

- artificial, 652
- mechanism of, 655-57, 667
- nature of, 654
- structure of, 656
- Semipolar bond, 99
- Sensitization of photographic plates, 1187
 - of sols, 1249
- Separation factor, of isotopes, 160
- Sharp series of spectra, 40, 41, 60, 66
- Silica, crystal structure, 388, 390
 - forms of, 471, 518
 - vitreous, 518-19
- Silicate glass, 519
- Silicates, structure of, 394-96
- Silk, structure of, 405-06
- Silver halides, photodecomposition 1186
 - iodide sols, 1241
 - silver chloride electrode, 942-43
- Simultaneous side reactions, 1085
 - Wegscheider's test for, 1086
- Single electrode potential, 937
- Singlet linkage, 106
- Smectic phase, 515
- Smith and Menzies apparatus, 447
- Soaps, *see* Colloidal electrolytes
- Sodium chloride, crystal structure, 367
 - lattice type, 385-87
 - sodium sulfate-water system, 804
 - water system, 775
 - energy levels, 61
 - sulfate-water system, 773-75, 786-87
- Sol, 1233; *see also* Colloidal systems
 - gel transition, 1258, 1262
- Solid, 340; *see also* Crystals
 - heat capacity of, 413-21
 - liquid systems, 742-70
 - rotation in, 422-24
 - solutions, 349, 379, 649, 651, 761-70
 - changes in, 766-69, 770
 - completely miscible, 761, 764-65
 - partially miscible, 766-69
 - of salts, 805-06
 - state, 340-424
 - vapor pressure, *see* Vapor pressure
- Solidus curve, 747, 761
- Solubility, calculation of, 743-44
 - curves, of solid, 843, 771
 - and Debye-Hückel theory, 962
 - determination by conductance, of gases, 693
 - coefficient of, 694
 - ideal, 702
 - in aqueous solutions, 699
 - mixtures, 693
 - product, 962, 970
 - determination, 972
 - and ionic strength, 963
 - and solubility, 970-72
- Solute, 625
- Solution pressure, 947
- Solutions, 625
 - dilute, 635, 644, 662, 671
 - and gas laws, 662-63
 - ideal, 675, 685
 - of gases in liquids, 693-703
 - ideal, 674-75
 - deviations from, 676-78; *see also* Raoult's law
 - nonideal, thermodynamics of, 683-90
 - pseudoideal, 678
- Solvation of ions, 897; *see also* Hydration
- Solvent, 625
 - associating, 681
 - dissociating, 680
 - effect in absorption spectra, 586
 - in acid-base equilibria, 975-77
 - on dipole moments, 549
 - in optical rotation, 605
- Solvolytic, 984
- Sørensen's pH scale, 982
- Sorption, 1194
- Space groups, 355
 - and crystal systems, 355
 - determination of, 358
 - lattice, 350-55
 - cubic, 353-54
 - determination of, 358
 - planes, 352
 - spacing of, 353
 - symmetry elements of, 354
 - types of, 54
 - and ionic radii, 385-91
- Specific catalysis, 1133
 - conductance, 889
 - dispersivity, 534
 - heat, *see* Heat capacity
 - magnetic susceptibility, 610

- tion rate, 1046
- raction, 529
- istance, 889
- otation, 596
- viscosity, 500
- pectra, absorption, *see* Absorption
- arc, 35
- atomic, 35-69
 - alternation in, 66
 - and wave mechanics, 77
 - Bohr theory of, 43-49
 - emission and absorption, 36, 44
 - excitation of, 35, 44
 - fine structure of, 41, 55, 58
 - constant, 57
 - of helium, 78-79
 - of hydrogen, 37-39, 49
 - hyperfine structure of, 92
 - multiplet structure of, 41, 62, 66, 68
 - origin of, 42
 - resonance lines in, 50
 - selection principle in, 58, 60
 - lines of, 37, 40, 41-42, 60, 66
 - and wave mechanics, 77
- ame, 35
- and isotopes, 147, 148
 - liquids, 580-88
- molecular, 563-80
 - band, 563, 564
 - continuous, 575, 1165
 - diffuse, 1166
- and nuclear spin, 93
- predissociation, 1166
- Raman, 576
- spark, 35
- X-ray, *see* X-rays
- Spectral multiplets, 62, 66-69
 - series, 37, 40-42
 - symbol, 69
 - terms, 36
- Spectroscopic displacement law, 36
- Speed of ions, 995, 908-10
- Spin, *see* Electron spin, and Atomic spin
- quantum number, electron, 61
 - of multiplet, 573
- of, 93
- tharoscope, 121
- ntaneous processes, 215
- and entropy change, 227-29
- Spreading coefficient, 485
 - of one liquid on another, 484-85, 1210
- Square terms of energy, 329, 331
- Stability of nuclei, 174
- Stalagmometer, 490
- Standard entropy of gases, 878
 - of ions, 1011-12
 - of liquids and solids, 869
- free energy, 827, 855
 - data, 857
 - of ions, 1010
- plane of crystal, 342
- potentials, 937, 944-47, 949-50
- states, of elements, 202
 - of gases, 301
 - of solvent, 684
 - of solute, 685
- Stannic oxide sol, 1242
- Stark-Einstein law of photochemical equivalence, 1159-60
- Stationary chains, 1083
 - state, 1079
 - photochemical, 1183
- Statistical methods, and entropy, 870-72
 - and free energy, 879-82
 - and heat capacity, 332-37
 - and ortho-para ratio, 95
 - weight, 95, 273
- Statistics, Maxwell-Boltzmann, 270-73
 - Bose-Einstein, 273
 - Fermi-Dirac, 274
- Steam distillation, 734
- Steric factor in reaction rate, 1098, 1102
- Stern-Gerlach experiment, 64
- Stoichiometric activity coefficient, 960
- Stokes's law, 12, 258, 261, 497-98, 897, 904
- Streaming potential, 1220, 1224
- Strong electrolytes, 892, 903
 - amplitude of X-rays, 359
 - of atom, *see* Atomic structure
 - of crystals, *see* Crystal structure
- Sublimation, 459, 467-68; *see also* Vapor pressure of solids
 - curve, 459
 - latent heat of, 459, 460, 462
- Sucrose, inversion of, 1052, 1132

- Sulfur, crystal structure of, 377
 dissociation of, 320
 liquid, 503-04
 paramagnetism of diatomic, 320, 618
 phase diagram of, 471-72
 and phase rule, 476
 trioxide formation, equilibrium, 838
- Super-acid solutions, 976
- Supercooling, 467, 647, 749
 avoidance of, 647
- Superheating, 434, 638
 avoidance of, 638-40
- Superlattice (or superstructure), 381
- Surface active substances, 1205
 area of adsorbent, 1218
 of catalyst, 1146
 energy, 481, 494
 free, 481, 1206
 total, 494
- films, 1210-15
 built up, 1218
 condensed, expanded and gaseous, 1212
 force-area curves of, 1212
 and gas laws, 1214
 multimolecular, 1214
 of soluble and insoluble substances, 1213
 unimolecular, 1211
 vapor pressure of, 1212
- force, 1213
- function in catalysis, 1146, 1147, 1150
- nonuniformity of, 1148
 and active centers, 1148-49
- oriented, catalytic activity, 1149
- orientation of molecules in, 1209
- phenomena, 1194-1280
- reactions, 1118-26
- tension, 481-96
 and adsorption, 1205-10
 and capillary rise, 486-87
 and critical temperature, 494-95, 526
 and density, 494-95, 526
 and electric charge, 1228
 and foam, 1251
 of mixtures, 496
 measurement of, 487-91
 and pressure differences, 486
 relationships, 487-90
- results, 491
 and spreading, 484
 and temperature, 491-93, 494-5
 and wetting, 484
- Susceptibility, magnetic, 6.
- Suspended transformation, .
- Suspensions, Brownian movement in, 255-58
 diffusion in, 259-63
 displacement of particles in, 259
- Swarm theory of liquid crystals, 51
- Swelling of gels, 1259, 1262
- Symbols, of atomic states, 69
 of molecular states, 572
- Symmetric functions, 79, 273
 vibrations, 570
- Symmetry, elements of, 341, 354
 number, 875
 closed, 231
 equilibrium in, 233-34
 isolated, 185
 open, 237
- Temperature, absolute, 192, 221
 coefficient of reaction rate, 1067
 of photochemical reactions, 11
 and kinetic energy, 252
 low, production of, 440-43
 measurement of, in calorimeter, 209
 reduced, 439
 thermodynamic scale of, 221-22
- Tensiometer, 490
- Termolecular reactions, 1063
 theory of, 1105-07
- Term, spectral, *see* Spectral term
- Ternary critical solution temperature, 797
 eutectic, 801
- Tetartohedral class, 346
- Thaw-melt method, 749
- Thermal analysis, 749, 762
 diffusion, separation of factors, 154-61
 dissociation, *see* Dissociation
 pressure in solution, 85
- Thermionic effect, 4
- work function, 33, 949
- Thermochemistry, 2

